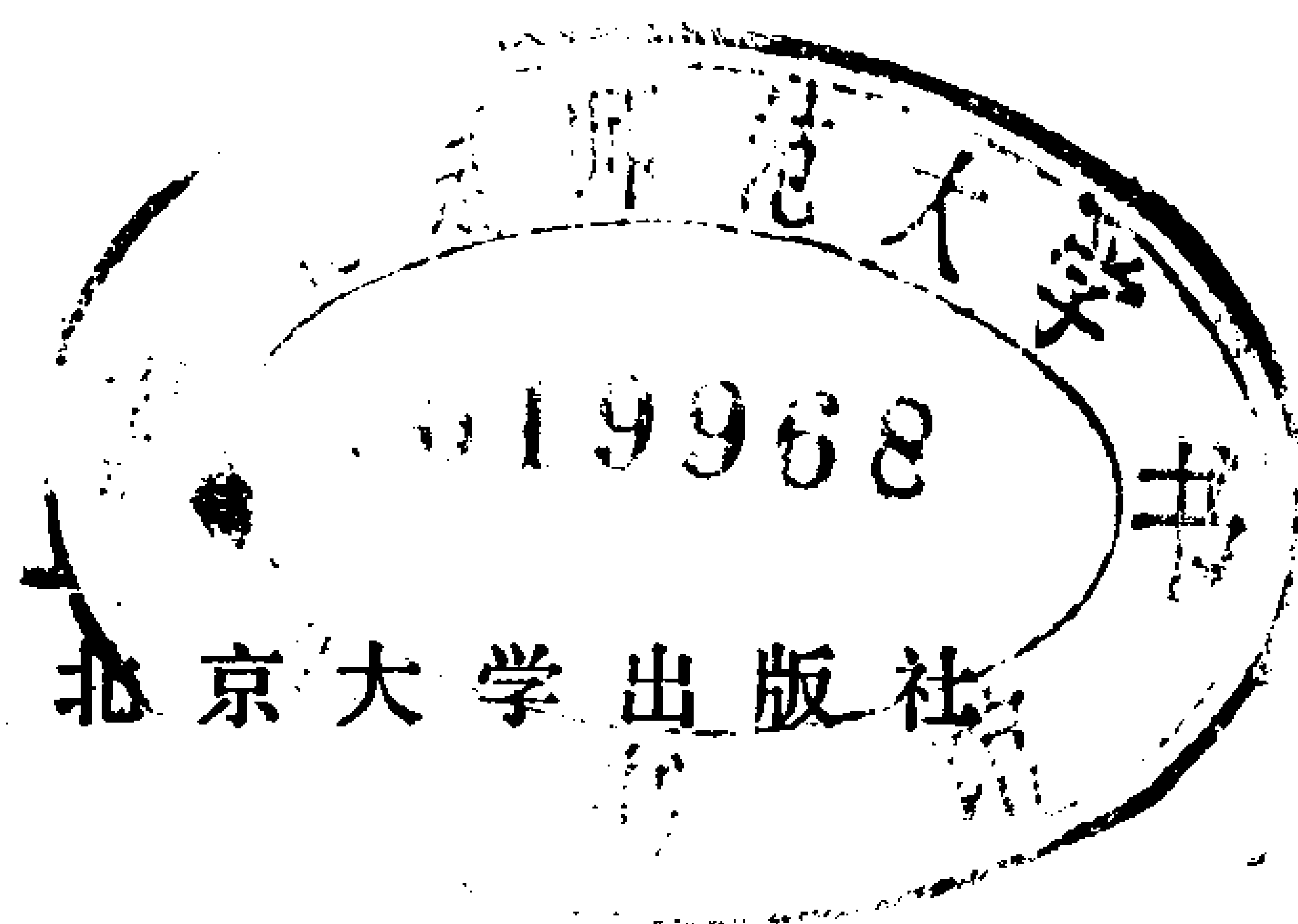
数学小丛书——智慧之花

（2）

# 邮票·自行车·果园·雨中行

丁石孙　主编




北京大学出版社

# 目　录

# PÓLYA 果园问题①②

T. T. Allen

## 1. 引言

在一个圆形的果园中，均匀地种植果树，问果树的树干长到多粗③，才能完全遮住果园中心的视线(Pólya 和 Szegö[6])？

设所有这些果树都是半径为 $r$ 的圆柱，这样，问题就简化为一个平面上关于圆的问题。设圆的中心（即果树的种植位置）的坐标为 $(x,y)$，$x,y$ 是整数，满足$\sqrt{x^2+y^2}\leqslant S$，$S$ 是果园的半径。射线是由原点向外的径直视线，第一个阻挡射线的圆沿这条射线是可见的。问题是要确定果树的半径 $\rho$，使得当 $r\geqslant\rho$ 时，在原点(即果园的中心)仅能看到果园中的果树；当 $r<\rho$ 时，至少有一条射线可穿过果园。显 然，$\rho$ 是 $S$ 的函数。

Pólya 的解法是基于 A. Speiser 的方法（参见上面所引的[6]，原始的解可见 Pólya[5]④. R. Honsbeigen[4] 的解法是基于 Minkowski 的凸体定理⑤. 但他们都未能求出 $\rho$ 的

---

① Pólya's Orchard problem, *Amer. Math. Monthly*, 93(1986), 98—104.

② Pólya的解答在文后。

③ 假定所有树的树干在生长过程中有同样的粗细。

④ 在本文最后，我们也译出了文献[6]中的解法。——译者注

⑤ 中译文见《数学译林》，1981年，第1期，79—84. ——译者注

确切值。他们只证明了：如果$S$是一整数，则

$$\frac{1}{\sqrt{S^2+1}}\leqslant\rho<\frac{1}{S}. \tag{1}$$

但是，$\rho$的确切值是多少？当$S$不是整数时，$\rho$的值又如何？

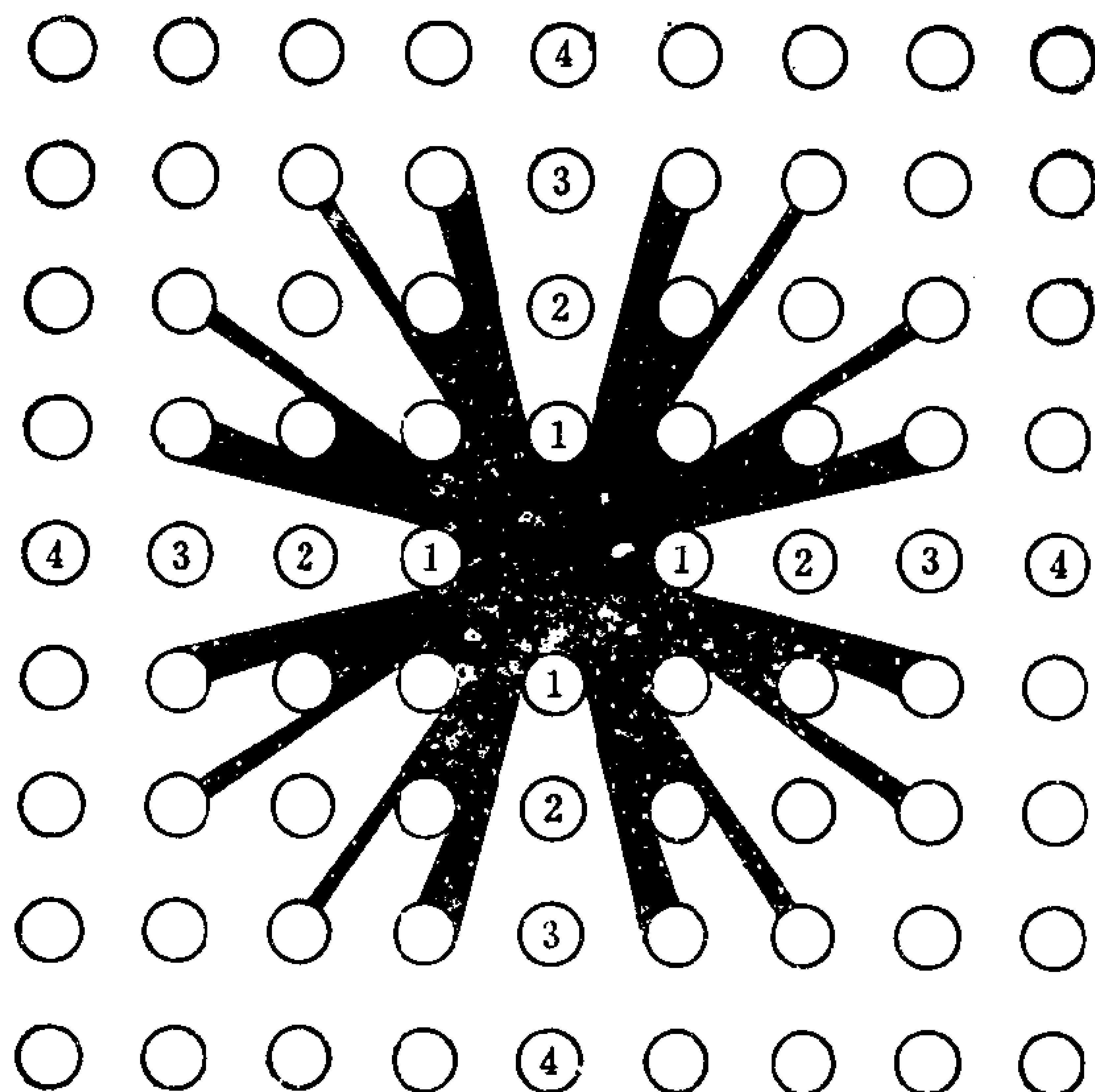



图1 $r=0.25$时射线按其在圆上的终点的划分图

下面计算格点$(x',y')$到射线（倾角为$\theta$）的矩离$r'$。注意到，$r'=h'\cos\theta$，$h'=y'-x'\operatorname{tg}\theta$，所以

$$r'=y'\cos\theta-x'\sin\theta.$$

如果射线通过格点$(x,y)$，则

$$\sin\theta=y/(x^2+y^2)^{1/2},\quad \cos\theta=x/(x^2+y^2)^{1/2},$$

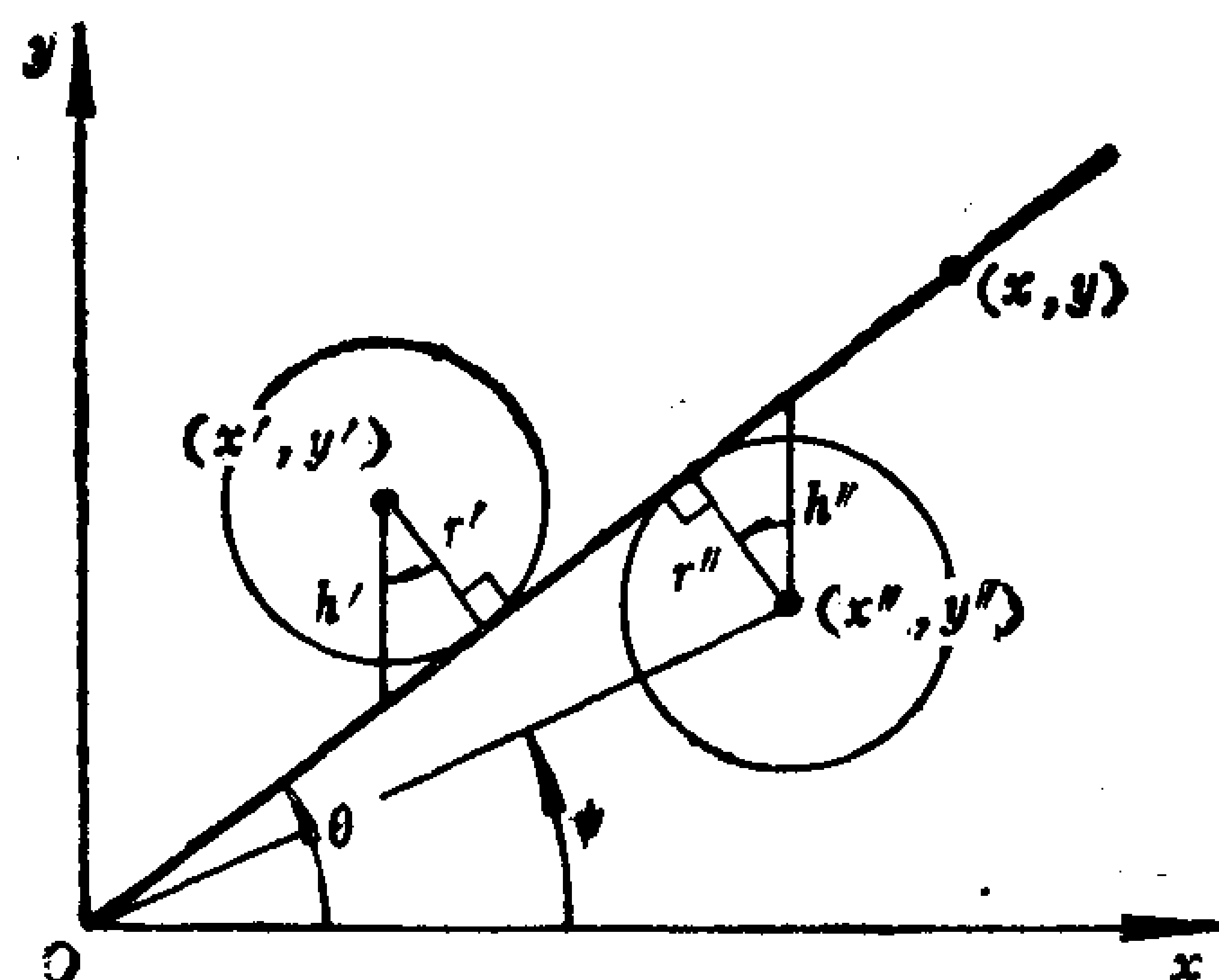



图 2　格点 $(x', y')$ 到倾角为 $\theta$ 的射线的距离

由此推出等式（2a）成立。同样可证，等式（2b）成立。对给定的 $r''$ 和 $\psi$ 计算 $\theta$（这里 $\psi = \operatorname{arc\,tg}\ (y''/x'')$），我们有

$$\theta - \psi = \arcsin\left(r''/\sqrt{x''^2 + y''^2}\right).$$

由此直接可得不等式（7d）。同样可推出不等式（7a），（7b），（7c）。

在本文中我们将用初等的方法证明：$\rho = 1/\sqrt{\lambda}$，$\lambda$ 是第一个大于 $S^2$ 且可以写成两个互素整数平方和的整数。如果 $S$ 是整数，则(1) 式左边的等号成立。我们还将提出两个与之有关的问题，以展示这个美丽的果园的其它性状。

## 2. 预备知识

考察图 1，我们有：

(1) 离原点最近的八个圆是对称的，因此，只要考虑第一象限内 $x \geqslant y \geqslant 0$，$(x, y) \neq (0,0)$ 的部分即可；

(2) 只有圆心是以互素的数为坐标的圆是可见的。例

如，圆心为（2,2）的圆完全被圆心为（1,1）的圆遮掩，这与圆的半径 $r$ 无关；

（3）在 $r=0$ 的极限情况下（这时所有的圆退化为点），只有由互素整数对组成的坐标点是可见的，它们是沿射线最先看到的点；

（4）在另一种极限情况 $r=\frac{1}{2}$ 时，这些圆相切，所以只有圆心在（1,0）和（1,1）（及它们在四个象限的对称点）的圆可见。

考虑通过格点 $(x,y)$ 的射线和射线两侧的格点 $(x',y')$，$(x'',y'')$，它们分别满足：

$$\frac{y'}{x'}>\frac{y}{x},\quad \frac{y''}{x''}<\frac{y}{x}.$$

从 $(x',y')$，$(x'',y'')$ 到射线的垂直距离分别为（参见图2）：

$$r'=(y'x-x'y)/(x^2+y^2)^{1/2},\tag{2a}$$

$$r''=(x''y-y''x)/(x^2+y^2)^{1/2}.\tag{2b}$$

式（2a）和（2b）中的分子可分别看作是点 $(x',y')$ 和点 $(x'',y'')$ 的函数，当 $x,y$ 互素时，它们都能取所有的正整数值（为什么？）。因此，最接近射线的点分别由

$$y'x-x'y=1,\tag{3a}$$

$$x''y-y''x=1\tag{3b}$$

给出，相应的极小距离是：

$$r'=r''=(x^2+y^2)^{-1/2}.\tag{4}$$

式（3a）和（3b）都是不定方程，若 $(\bar{x}',\bar{y}')$ 和 $(\bar{x}'',\bar{y}'')$ 分别是它们的特解，则它们的通解分别为 $(kx+\bar{x}',ky+\bar{y}')$

和 $(kx+\bar{x}'', ky+\bar{y}'')$ ，这里 $k$ 为任意整数。这些解分布在与射线等距离的两条平行线上，距离就是由(4) 式给出的极小距离。为了讨论可见性，我们需要这样的解，它所确定的点距原点较 $(x,y)$ 近，且与 $(x,y)$ 在同一象限中，即

$$0\leqslant x'\leqslant x, \tag{5a}$$

$$0\leqslant y''<y。 \tag{5b}$$

在(5a)给出的区间内恰好存在一组坐标对 $(x',y')$ 满足(3a)，这是因为在(3a)的通解 $(kx+\bar{x}', ky+\bar{y}')$ 中恰有一个解，使得 $kx+\bar{x}'$ 落在每一个长为 $x$ 的半开区间中。同样，必有满足(3b)的唯一坐标对 $(x'',y'')$，使得 $y''$ 在(5b)所给的区间中。

上述论证表明：圆心由(3)，(5)确定，半径由(4)确定的两个圆与通过点 $(x,y)$ 的射线相切。同样，射线在与以 $(x,y)$ 为圆心的圆上一点相碰前，先碰上了上面两个圆与射线的切点。因为由 Pythagoras 定理知，这两个切点沿射线到原点的距离分别为

$$(x''^2+y''^2-1/(x^2+y^2))^{1/2}$$

和

$$(x'^2+y'^2-1/(x^2+y^2))^{1/2}①。$$

参看图 1 可知，这种距离的最小值出现在点 $(x,y)=(2,1)$，$(x'',y'')=(1,0)$，所以，这些距离总是不小于 $[1^2+0^2-1/(2^2+1^2)]^{1/2}=\frac{2}{\sqrt{5}}$。而由式(4)知，这些相切圆的半径一定不大于 $(2^2+1^2)^{-1/2}=\frac{1}{\sqrt{5}}$。

---

① 容易算出，从原点到射线与以 $(x,y)$ 为心的圆的第一个交点之间的距离为 $(x^2+y^2)^{1/2}-(x^2+y^2)^{-1/2}$，显见，它比这两个数都大。——校注

最后，我们得到了这样的结论：对互素的整数对 $x,y$，当 $r<(x^2+y^2)^{-1/2}$ 时，以点 $(x,y)$ 为心半径为 $r$ 的圆至少沿一条射线是可见的；而当 $r\geqslant(x^2+y^2)^{-1/2}$ 时，这个圆则被完全遮掩。

### 3. 果园问题

首先，假定这个果园是处处可伸展到无穷远的。我们围绕在原点的观察哨扎一个半径为 $S$ 的篱笆，并把这个篱笆看作是果园里的一个圆。问当果树的半径 $(r=\rho)$ 是多少时，正好遮掩住篱笆外的树？

由(4)知，一棵树要被遮掩住，只要半径 $r$ 等于从原点到这棵果树圆心距离的倒数。因此，在果树生长过程中，篱笆外最后被遮住的一棵树① 是长得最靠近篱笆的一棵，因为所有离原点更远的树已在树干半径 $r$ 较小时被遮住了。故从原点到篱笆外最近的一棵树的圆心的距离的倒数就是所要求的 $\rho$。证毕。

上述论证并不依赖于篱笆外是否真的长有树，因为决定可见性的仅是篱笆内的树。

如果 $S$ 是一个整数，那么 $S^2+1$ 是第一个大于 $S^2$ 的整数。在点 $(S,1)$ 处总有一棵树，它到原点的距离是 $(S^2+1)^{1/2}$。能看到篱笆外面的临界半径 $r=\rho=(S^2+1)^{-1/2}$。当然，还可能有别的树和原点有相同的临界距离 $(S^2+1)^{1/2}$。例如，当 $S=50$ 时，在点 $(50,1)$，$(49,10)$ 及这两点在四个象限的其它14个对称点处的树恰好在半径 $r=\rho=\frac{1}{\sqrt{2501}}$ 时同时消失于视

① 实际上是一些离原点等距的树。——译者注

野。

要指出的是，这里得到的表示式 $\rho=(S^2+1)^{-1/2}$ 与(1)式的左边的等式相同，故当 $S$ 是整数时，(1)式中左边的等号成立。

如果 $S$ 不一定是整数，那么这个问题的一个更有启发性的表述是："$\rho$ 和 $S$ 将是怎样一个函数关系?"同样，设 $x, y$ 是任意互素整数，$x^2+y^2=\lambda$，根据事实本身来看，$\lambda$ 是一种特殊的整数——这种整数至少可以以一种方式表示成两个互素整数的平方和。假定 $\lambda_i$ 和 $\lambda_{i+1}$ 是两个相邻的这种整数，

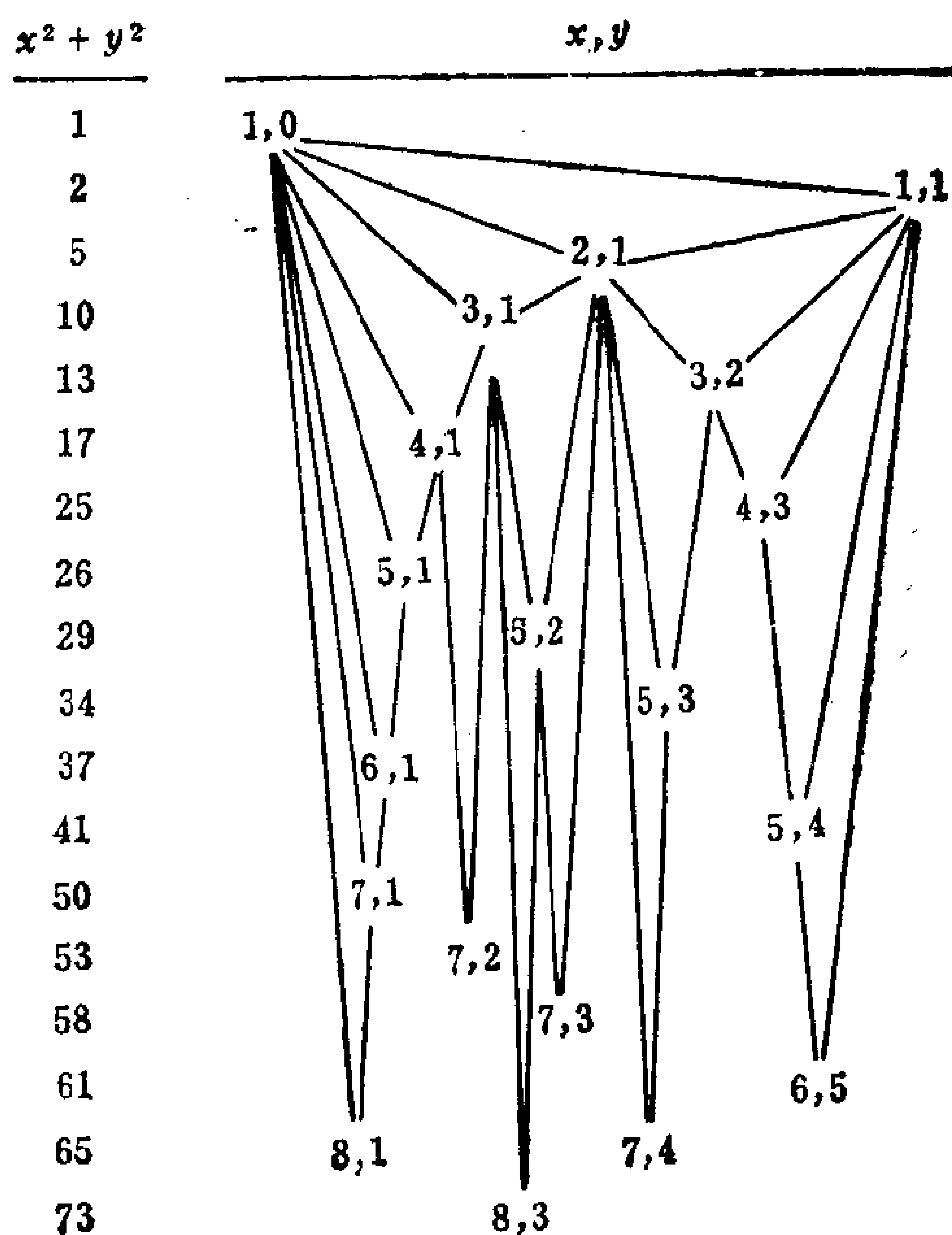



图 3　随着半径的增长树被遮掩的情况

$\lambda_i<\lambda_{i+1}$，那么当$\sqrt{\lambda_i}\leqslant S<\sqrt{\lambda_{i+1}}$时，相应于这些$S$的临界半径$\rho$一定是常数，且$\rho=\frac{1}{\sqrt{\lambda_{i+1}}}$，因为$\sqrt{\lambda_{i+1}}$是篱笆外的第一棵树的树心到原点的距离①。另一种证法是：当$1/\sqrt{\lambda_{i+1}}\leqslant r<1/\sqrt{\lambda_i}$时，可见的最远的树的圆心距离原点为$\sqrt{\lambda_i}$，因为距原点$\sqrt{\lambda_{i+1}}$的树在半径$r=1/\sqrt{\lambda_{i+1}}$时已被遮掩。

图3按$\lambda$的递增顺序列出了$\lambda=x^2+y^2$的互素数对。图中这些连线是表示每一棵树恰好可被离原点较近的另外两棵树遮掩②。65和73是两个相邻的$\lambda$值。有趣的是65是第一个可以用两种方法表示成互素数对平方和的数。因为$1/\sqrt{73}\leqslant r<1/\sqrt{65}$，所以，可见的最远的树在(8,1)和(7,4)（以及它们在四个象限的对称点）处。如果$r\geqslant 1/\sqrt{65}$，它们就同时消失于视野。假如我们以区间$\sqrt{61}\leqslant S<\sqrt{65}$内的任一值$S$为半径围一个篱笆，那么可见的临界半径是$r=\rho=1/\sqrt{65}$。

## 4　连分数

如果我们始终保持沿一条射线注视生长着的树，我们将看到哪些树呢？

我们把这条射线的斜率$\mathrm{tg}\theta$展为连分数，如果$p_n/q_n$是它的$n$阶渐近分数，那么我们知道，对于射线的斜率$\mathrm{tg}\theta$来说，$p_n/q_n$是所有分母小于或等于$q_n$的分数中好的有理逼近，

① 已经假定树心，即种树处的格点$(x,y)$中的$x,y$是互素的。——译注

② 例如，在$(x,y)=(3,1)$处的树被在$(x',y')=(2,1)$，$(x'',y'')=(1,0)$处的树遮住，即满足(3a)，(3b)的两个解。——校注

也就是说(见 Hardy 和 Wright [3] 定理 181①)，如果 $n>1$，$0<q\leqslant q_n$ 且 $p/q\neq p_n/q_n$，则

$$|p_n-q_n\mathrm{tg}\theta|<|p-q\mathrm{tg}\theta|. \tag{6}$$

用几何语言来描述，式(6)就是说：从点$(q_n,p_n)$沿竖直方向（即 $y$ 轴方向）到这射线的距离比任何所说的点$(q,p)$到这射线的距离要短．这些沿竖直方向的距离就是图 2 中的那些 $h$，我们把在射线上方的记为 $h'$，在射线下方的记为 $h''$．(6)式的两端同乘 $\cos\theta$，就给出了这些点到这射线的垂直距离 $r$ 的表示式．因为 $\cos\theta$ 在第一象限为正，这样，定理的结论就转化为关于这些垂直距离之间的关系：点$(q_n,p_n)$比这些点$(q,p)$中的任意一点离射线更近．

我们知道 $q_n<q_{n+1}$，所以，由该定理推出：对所有的 $n$，点$(q_{n+1},p_{n+1})$比点$(q_n,p_n)$距射线近．如同前面的讨论一样，可知：圆心在点

$$\cdots,(q_{n+1},p_{n+1}),(q_n,p_n),(q_{n-1},p_{n-1}),\cdots$$

的圆就是当 $n$ 递增时沿这射线依次可见的那些圆．

若 $\mathrm{tg}\theta$ 是有理数，则渐近分数的个数和可见的树的数目都是有限的，因为当 $r=0$ 时，我们只看到一个格点。这里要指出的是，所展成的有限连分数的正确形式必需是其最后一项系数（即最后一个部分商）一定要取大于 1，否则，由简单的计算就可表明，连分数的倒数第二、三个渐近分数所给出的两个点是等距地位于射线两侧，而仅有接近原点的那个圆是沿这条射线可见的。例如，若 21/16 的连分数展开式取

① 关于连分数的知识可参看：华罗庚著《数论导引》第十章，及奥尔德斯的《连分数》(北京大学出版社)．——校者注

为[1,3,4,1]，它的渐近分数是1/1,4/3,17/13,及21/16.但沿 $\text{tg}\theta=21/16$ 只能看到以(1,1),(3,4)及(16,21)为圆心的圆，所以我们应取连分数展开式21/16=[1,3,5]。

若 $\text{tg}\theta$ 是无理数，它的渐近分数和可见树的数目都是无限的。例如，当斜率 $\text{tg}\theta=(\sqrt{5}+1)/2$ 时，对于 $r\leqslant 1/2$，可见圆序列可以看作是递减的半径 $r$ 的函数，这是一个 Fibonacci 序列(1,1),(1,2),(2,3),(3,5),(5,8),(8,13)等等。应该指出的是，除非这些果树交叠、即 $r\geqslant 2((\sqrt{5}+1)^2+2^2)^{-1/2}\approx 0.53$，不然，在点(0,1)处的树沿这条射线是不可见的(为什么?)。

## 5. 单棵树的可见性

设 $x,y$ 是任意互素数对，满足 $x\geqslant y\geqslant 0$，$(x,y)\neq(0,0)$。我们断言：位于这点处的圆当 $r\leqslant\frac{1}{2}$ 时，在下面给出的不等式的交集所确定的角域中是可见的（角 $\theta$ 表示和正向 $x$ 轴的夹角）：

$$\theta\leqslant\text{arc tg}\frac{y}{x}+\text{arc sin}\frac{r}{(x^2+y^2)^{1/2}},0\leqslant r<r^+,\quad(7a)$$

$$\theta<\text{arc tg}\frac{y'}{x'}-\text{arc sin}\frac{r}{(x^2+y^2)^{1/2}},r^+\leqslant r<r^0\quad(7b)$$

$$\theta\geqslant\text{arc tg}\frac{y}{x}-\text{arc sin}\frac{r}{(x^2+y^2)^{1/2}},0\leqslant r<r^-,\quad(7c)$$

$$\theta>\text{arc tg}\frac{y''}{x''}+\text{arc sin}\frac{r}{(x^2+y^2)^{1/2}},\quad r^-\leqslant r<r^0,\quad(7d)$$

其中$(x',y')$，$(x'',y'')$满足(3)式和(5)式，

$$r^0=(x^2+y^2)^{-\frac{1}{2}},$$

$$r^+=[(x+x')^2+(y+y')^2]^{-\frac{1}{2}},$$

$$r^-=[(x+x'')^2+(y+y'')^2]^{-\frac{1}{2}}.$$

§2的证明不能得到这一结果。我们需要证明，圆心在$(x',y')$和$(x'',y'')$的圆不但决定了完全遮掩以$(x,y)$为心的圆的临界半径$r^0$，而且还决定了上面所断言的部分遮掩这个圆的所有角度。

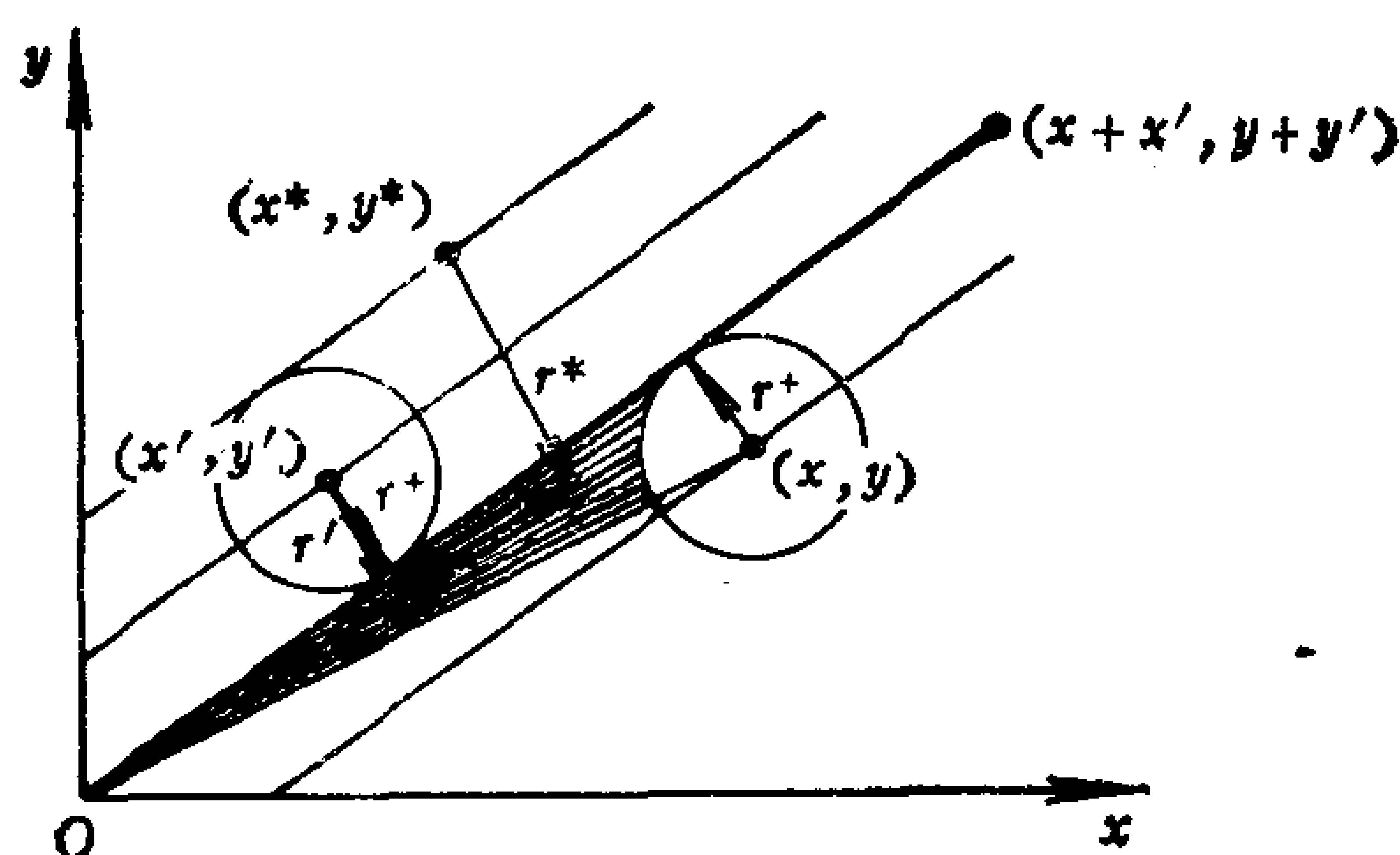



图4　圆心在$(x,y)$和$(x',y')$（满足(3a)与(5a)），半径为$r=r^+$时两圆之间的关系

从图4可以看出，圆心在$(x,y)$和$(x',y')$的圆，当它们的半径$r=r^+$时，与通过$(x+x',y+y')$的射线相切。超过这个临界半径时，前一个圆就从上方侵占了后一个圆的可见性。圆心在$(x^*,y^*)$ $(y^*/x^*>y'/x')$的圆是与此无关的。

考虑图4给出的图形，应用公式(2)，我们得到：$(x,y)$

和$(x',y')$以极小距离$r^+$等距地分布在通过$(x+x',y+y')$的射线的两侧，因此，当$r<r^+$时，圆心在$(x,y)$的圆的上半部都是可见的，(7a)给出了可见这个圆的射线的角度的上界(在图2的说明中，给出了$(x,y)$函数和$r$的形式)。然而，当$r^+\leqslant r<r^0$时，与圆心在$(x',y')$的圆相切的下方的射线是和图4中由阴影表示的楔形部分中的射线是一致的，因此，以$(x,y)$为心的圆的上部是部分地被遮掩的。点$(x',y')$到这楔形中所有射线的距离都小于$2r^+$。满足条件$y^*/x^*>y/x$及$0\leqslant x^+<x$的其它点$(x^*,y^*)$到这楔形中的射线的距离都大于

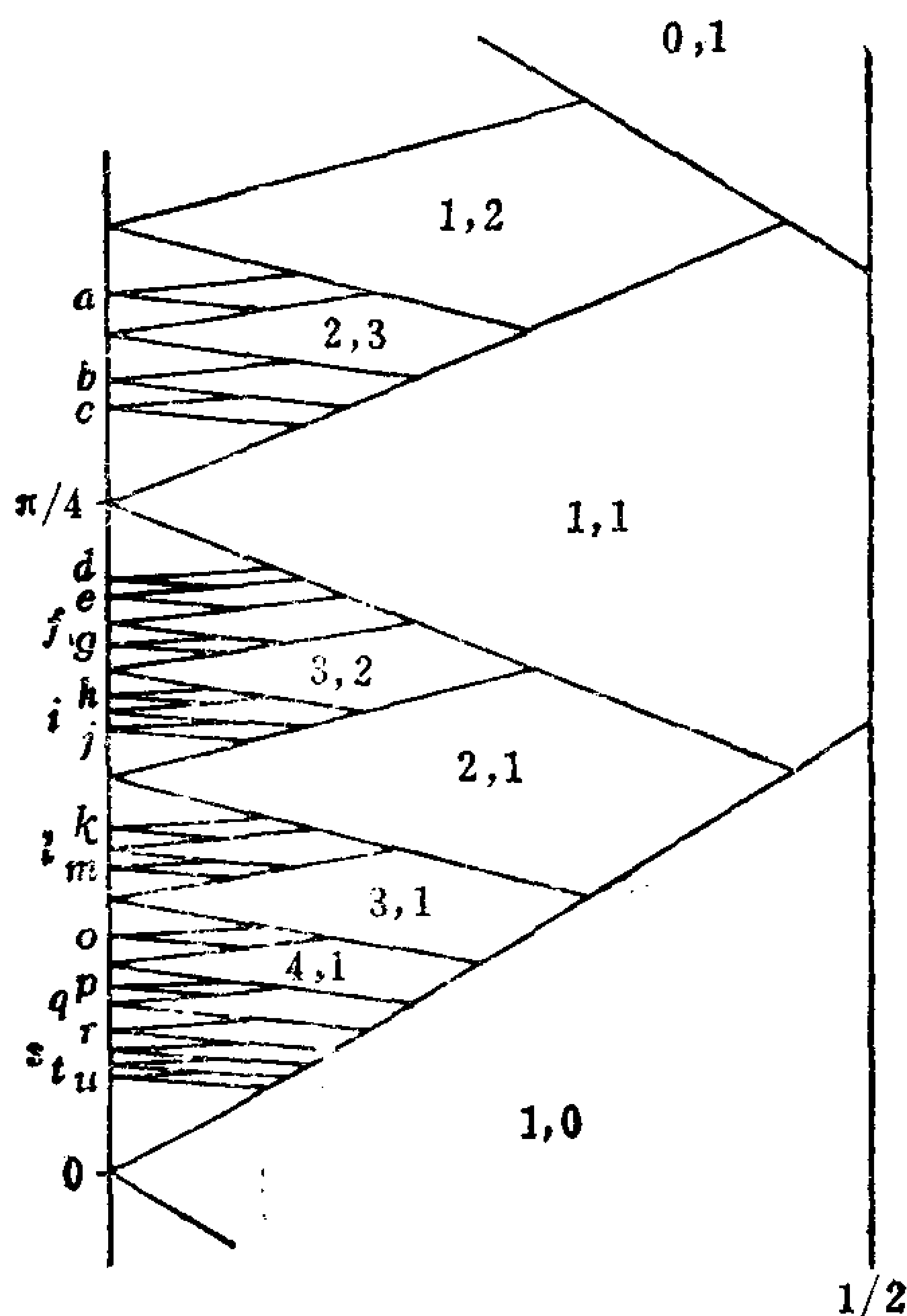



**图5　角的区间(在这范围内不同的圆是可见的)是圆的半径的函数**

$2r^+$，因为所有格点与通过$(x'+x, y'+y)$的射线的距离是$r^+$的整数倍。在区间$[0, x)$内只有点$(x', y')$在这射线的上方且恰好相距$r^+$，因此，当$r^+ \leqslant r < r^0$时，(7b)就是给出这个可见角域的上界部分的充分必要条件。同样的讨论可应用于以$(x'', y'')$为心的圆从下方部遮掩的情形，而(7c)，(7d)就给出了这个可见角域的下界部分。

在图5中，纵坐标旁边的字母相应地表示圆心所在的位置：

$a(3,5)$，$b(3,4)$，$c(4,5)$，$d(6,5)$，$e(5,4)$，
$f(4,3)$，$g(7,5)$，$h(8,5)$，$i(5,3)$，$j(7,4)$，
$k(7,3)$，$l(5,2)$，$m(8,3)$，$n(3,1)$，$o(7,2)$，
$p(9,2)$，$q(5,1)$，$r(6,1)$，$s(7,1)$，$t(8,1)$，
$u(9,1)$。

我们对足够多的$(x, y)$，把相应于它们的由(7a)到(7d)所确定的那些角域转化成图中所示的$r-\theta$平面上的嵌砖形区域。显然，每一个互素数对$(x, y)$将在这$r-\theta$平面上占有一块区域，当$x, y \to \infty$时，它们将充满整个$r-\theta$平面。

6．附录

代替这种圆形格，我们可以讨论三角形格的可见性问题，这个问题是由双神经原(例如，协调心脏跳动的神经原)的一个简单模型引出的。这种三角形格也是某些电路(例如，保持电视机的图像稳定的电路)的一个颇佳的模型。解决这个问题的方法和本文给出的很相似。例如，一个类似图5的图描述了如何依据两个振荡器之间的相互作用强度及它们的自然周期之比，来使得一个振荡器在另一个振荡器周期的有

理倍数上同步，并指出了这种同步追踪能如何有效地抗干扰。

## 参 考 文 献


[1] T. T. Allen, On the arithmetic of phase locking: Coupled neurons as a lattice on $R^2$, *Phys.* D, 6 (1983), 305—320.

[2] ______, Complicated, but rational, phase locking responses of a simple time-base oscillator, IEEE Trans. Circuits and Systems, 30 (1983), 627—632.

[3] G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, Clarendon Press,Oxford, 1979.

[4] R. Honsberger, Mathematical Gems I, Dolciani Mathematical Expositions, no.1,Chap.4, Mathematical Association of America, Washington, DC, 1973.

[5] G. Pólya,Zahlentheoretisches und wahrscheinlichkeitstheoretisches uber die sichtweite in walde,*Arch Math. und Phys.*, 27, Series 2 (1918), 135—142.

[6] G. Pólya and G. Szegö, Problems and Theorems in Analysis, Vol.2, Chap.5, Problem 239,Springer-Verlag, New York, 1976.


## PÓLYA 的解答

下面是文献[6]中关于果园问题的解法，是该书的第239题。所用符号的意义和本文大致相同。

**问题** 设$S$是正整数。设每一格点$(p,q)$是半径为$r$的圆的圆心，且满足不等式$1\leqslant p^2+q^2\leqslant S^2$。如果$r$充分小，则存在从(0,0)到无穷远的射线，这些射线与上面所述的小圆不相碰(此时，果园称为是透亮的)；当$r=1/2$时，这些圆

相切，所以，$r$ 足够大时，这样的射线将不存在。设 $r=\rho$ 是划分这两种情形的临界值(即使果园是透亮的极限值)，那么我们有

$$1/\sqrt{S^2+1}\leqslant\rho<1/S.$$

**解答**(G. Pólya: Arch Math. Phys. Ser. 3, 27, 135 (1918); method of proof given here by A. Speiser).

我们说格点 $(p,q)$ 是本原格点，如果它是从原点可见的(简称可见的)，即在联结 $(0,0)$ 和 $(p,q)$ 的线段上没有其它的格点.容易证明:格点 $(p,q)$ 是本原的充要条件是 $p,q$ 互素(为什么？)。若 $pv-qu=1$，则格点 $(p,q)$，$(u,v)$ 都是本原的，且它们由一个面积为 1 的平行四边形相连(另外两个顶点是 $(0,0)$ 和 $(p+u,q+v)$)。我们称 $(u,v)$ 是 $(p,q)$ 的左邻点，$(p,q)$ 是 $(u,v)$ 的右邻点，以及把从原点 $(0,0)$ 出发的那条对角线称为是这个相连平行四边形的对角线。如果 $(p,q)$，$(u,v)$ 的相连平行四边形的对角线长为 $d$，则 $(p,q)$，$(u,v)$ 到对角线的距离为 $1/d$(为什么)。每个本原格点有无穷多个左邻点，它们均匀地分布在一条直线上①(为什么？)。

(1) $(1,0)$ 和 $(S-1,1)$ 是相邻的，它们的平行四边形的对角线长为 $\sqrt{S^2+1}$。如果这条对角线的延长线和一个 $\rho$-圆相交，那么只需要考虑以 $(1,0)$ 和 $(S-1,1)$ 为心的圆，故 $\rho\geqslant 1/\sqrt{S^2+1}$。

(2) 对 $x^2+y^2\leqslant S^2$ 内的任意本原格点，决定它在该圆内最远的左邻点 $(p',q')$，即 $(p'+p,q'+q)$ 已在圆 $x^2+y^2\leqslant$

① 请读者自己写出给定的本原格点 $(p,q)$ 的所有左邻点，并证明相邻的两个左邻点之间的距离为 $\sqrt{p^2+q^2}$。——校注

$S^c$外。在同样的意义下，设$(p'',q'')$是$(p',q')$的最远的左邻点，$(p''',q''')$是$(p'',q'')$的最远的左邻点，等等。这样做若干步，比如说$n$步之后，我们就得到具有这样性质的$(p^{(n)},q^{(n)})$：$(p,q)$与$(p',q')$的相连平行四边形，$(p',q')$与$(p'',q'')$的相连平行四边形，…，以及$(p^{(n-1)},q^{(n-1)})$与$(p^{(n)},q^{(n)})$的相连的平行四边形合在一起，完全覆盖了圆$x^2+y^2\leqslant 1$。$(p,q)$和$(p',q')$的相连平行四边形的对角线长大于$S$，并且$(p',q')$和$(p,q)$到这对角线的距离小于$1/S$。因此，从原点$(0,0)$出发的每条射线都被以$(p,q)$，$(p',q')$，…，$(p^{(n)},q^{(n)})$为圆心，$1/S$为半径的圆中的某一个所遮住，而这些对角线实际上是被两个圆所遮住，故有$\rho<1/S$。

（王明舟译，潘承彪校）

# 自行车问题[①]

D.E.DAYKIN

下面的老问题是很有启迪性的：

设静置自行车的脚蹬曲柄处于铅直位置，如果底部的脚蹬被轻轻地向后推动，问自行车将向哪个方向运动？

力的作用所引起的脚蹬转动，与我们骑车时的情形是一致的，因而大多数人会回答：“当然向前！”但只要他们推车到大街上一试，就会发现自行车是向后运动的．这就有趣了．人们引入摩擦力、力偶、牛顿定律，甚至还要计算齿轮的齿数，想对事情作些解释．其实道理如下所述，简单明了．

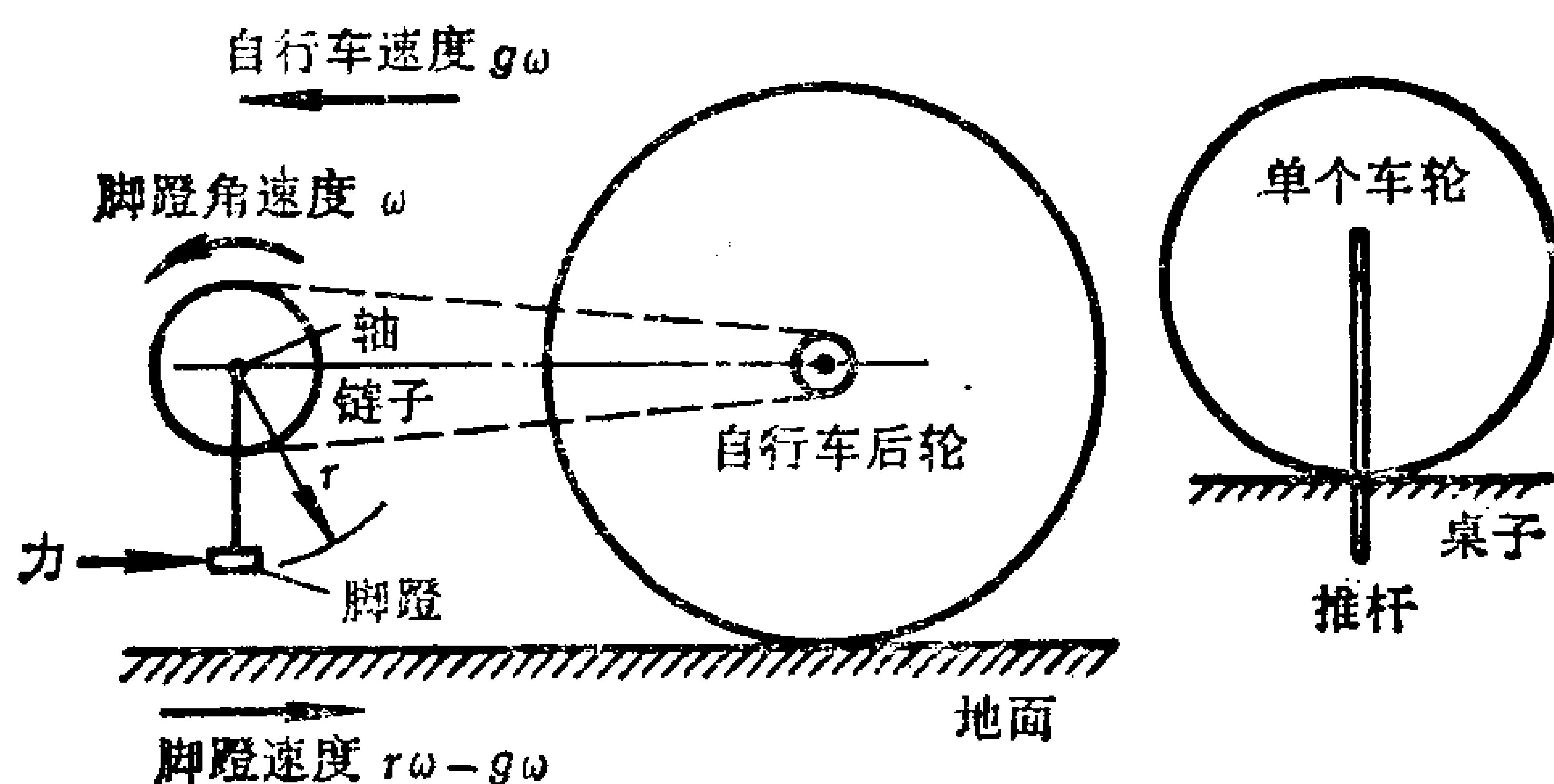



图 1

① The bicycle problem, *Math. Magazine*, 45 (1972), p 1.

如图 1 所示，设力引起脚蹬绕中轴旋转的角速度为 $\omega$。如果自行车的传动比为 $g$，那么自行车相对于地面前进的速度为$g\omega$。此外，如果 $r$ 是脚蹬的旋转半径，由于脚蹬是在它的最低位置，所以脚蹬相对于地面向后的速度是$r\omega-g\omega$。既然脚蹬是力的作用点，其运动不会与力反向。换句话说，如果脚蹬运动且$(r-g)\omega\geqslant0$，那么运动就是向后的。因此或者$r>g$且$\omega\geqslant0$，或者$r<g$且$\omega\leqslant0$，或者$r=g$。但前面已指出自行车的速度是 $g\omega$，并且 $g>0$，所以当 $\omega>0$时，自行车向前运动；当$\omega<0$时，自行车向后运动。普通的自行车$r<g$，自行车向后运动。注意到如果$r=g$，自行车将在不稳定的平衡状态中保持静止，因为脚蹬的速度为零，力也不能做功。

为进一步阐明这种情形，我们考虑在一张桌子上放上单个的车轮，使得它与桌子的边缘在一个铅直平面上，假设车轮有一个固定的径向长推杆，它可以从车轮的中心垂直悬挂到桌子水平面下方。如果我们推动长杆使车轮转动，轮子与桌子之间的接触点将是瞬时静止的，所以轮子转动的方向取决于推这长杆的位置是在桌面之上，还是在桌面之下。

（刘 勇译，陈一梦校）

**编者按**　这两篇文章是讨论人在雨中行走时尽量减少挨淋的问题。第一篇文章出现了一些技术性的错误，第二篇文章作了纠正。现在，我们仍完全按原样译出，目的在于让读者更清楚地看到，为了把一个实际问题抽象成数学问题，应如何抓住问题的本质，使建立的数学模型与实际问题尽量没有原则差异。有兴趣的读者，在读完了第一篇文章之后，可以尝试一下，去寻找其中的错误。

# 雨　中　行[①]

M. Deakin

**摘要**　考虑一个人在雨中走(或跑)的简化的模型。如果想在走过一段固定的距离之后，身上被淋湿的部分尽可能地少，那末他应以什么速度前进呢？一般说来，应该是跑得越快越好，但并不总是这样。

## 1. 问题

下着连绵的雨，设雨丝是平行的。一个人必须在雨中从$A$走(或跑)到$B$。他没有带伞，因而要寻求前进的速度($u$)，

① Walking in the rain, *Math. Magazine*, 45 (1972), 246—253.

使得落在身上的总雨量最少。我们来求 $u$ 的最佳值。

先建立这一问题的数学模型如下。设 $\boldsymbol{i}$ 是方向 $\overrightarrow{AB}$ 的单位向量，$\boldsymbol{k}$ 是方向朝上的单位向量，并令 $\boldsymbol{j}=\boldsymbol{k}\times\boldsymbol{i}$①。假设雨点的大小和形状都是一样的，落下的速度是 $V_T$，并被速度是 $V_T(w\boldsymbol{i}+W\boldsymbol{j})$ 的水平方向的风吹动。因此雨速是

$$\boldsymbol{V}=V_T(w\boldsymbol{i}+W\boldsymbol{j}-\boldsymbol{k}). \tag{1}$$

利用 $V_T$，人的速度 $u$ 可以表为

$$u=xV_T, \tag{2}$$

因此相对于此人的雨速是

$$\boldsymbol{V}_{rel}=V_T\{(w-x)\boldsymbol{i}+W\boldsymbol{j}-\boldsymbol{k}\}. \tag{3}$$

我们把人设想为一个长方体，6 面中有 3 面受到雨淋：前面(或后背)，一侧(取右侧)及顶部。设它们的面积分别是 $A$，$\eta A$ 及 $\varepsilon A$。

## 2. 淋湿函数

淋在头顶上的雨量正比于面积 $\varepsilon A$ 乘上这块表面的法向量 $\boldsymbol{k}$ 与雨的相对于人的速度方向的夹角的余弦。余弦值为

$$\frac{1}{\sqrt{(w-x)^2+W^2+1}}.$$

因此，落在顶部的雨量正比于

$$\frac{\varepsilon A}{\sqrt{(w-x)^2+W^2+1}}.$$

① 这是向量积，$\{\boldsymbol{i},\boldsymbol{j},\boldsymbol{k}\}$ 构成右手系。

类似可得，落在右侧表面上的雨量正比于

$$\frac{\eta AW}{\sqrt{(w-x)^2+W^2+1}}.$$

落在背部或前面的雨量的计算稍复杂些，它正比于

$$\frac{A|w-x|}{\sqrt{(w-x)^2+W^2+1}},$$

取绝对值是因为考虑到，不管雨是从前面还是从后面袭来，人被淋湿的情形都是相同的这一事实。

于是，落在人身上的总雨量正比于

$$\frac{\varphi+|w-x|}{\sqrt{(w-x)^2+n^2}},$$

其中

$$\varphi=\varepsilon+\eta W, \tag{4}$$

$$n^2=1+W^2. \tag{5}$$

落在人身上的总雨量正比例于单位时间落在他身上的雨量，且与人的速度成反比——即，与

$$F(x)=\frac{\varphi+|w-x|}{x\sqrt{n^2+(w-x)^2}} \tag{6}$$

成正比。我们称 $F(x)$ 为“淋湿函数”，下面求 $F(x)$ 的最小值。

### 3. 走进雨中

如果雨向一个静止不动的人的前面袭来，则 $w$ 是负的。我们发现，在这种情形时，重新定义 $w$ 使其为正可以更方便些。这略微改变了 $F$ 的表达式，现在

$$F(x)=\frac{\varphi+(w-x)}{x\sqrt{n^2+(w-x)^2}}. \tag{7}$$

容易证明，对于$x>0$(这是我们感兴趣的区 域)，$F(x)$是单调递减函数。因此当$x$取可能的最大值时，$F(x)$最小。从而，进入雨中的最好的策略是跑得尽可能地快。

### 4. 走出雨区

$w>0$的情形导致直接由(6)式给出的函数$F(x)$。$F'(x)$在$x=w$处不连续。按此速度前进，不管是人的背部还是前部，都淋不到雨。比这个速度快，人就超过雨点，从而前部挨淋。比这个速度慢，人的背部仍挨雨淋。

容易求得

$$F(w)=\frac{\varphi}{nw}, \tag{8}$$

$$F'(w-)=-\frac{1}{nw^2}(w+\varphi), \tag{9}$$

$$F'(w+)=\frac{1}{nw^2}(w-\varphi). \tag{10}$$

注意到$F'(w-)$永远是负的。因此，如果$F'(w+)>0$，即$w>\varphi$，则$x=w$是极小值点。$w<\varphi$和$w>\varphi$的情形将在下面分别讨论，且分别记为情形1和情形2。

### 5. 当$x$较小时$F(x)$的变化

当$x<w$时，

$$F(x)=\frac{\varphi+w-x}{x\sqrt{n^2+(w-x)^2}}.$$

我们要弄清楚，这个函数在$(0,w)$中是否有转向点(turning point)。微分$F(x)$并令$F'(x)=0$得方程

$$x^3-(3w+2\varphi)x^2+3w(\varphi+w)x-(n^2+w^2)(\varphi+w)=0. \tag{11}$$

由 Descartes 符号法则①，这个三次方程或者有 1 个或者有 3 个正根且没有负根。令 $y=w-x$，得

$$y^3+2\varphi y^2-w\varphi y+n^2(\varphi+w)=0. \tag{12}$$

由 Descartes 符号法则，它有 1 个负根，而且或者有 2 个正根或者没有正根。这两个正根，如果有的话，出现在我们感兴趣的区域中；这当且仅当方程(12)有 3 个根(指实根——译者注)时才能发生。应用三次方程的判别式条件，可以发现当

$$\left(\frac{\varphi}{3}\right)^3\left(\frac{4\varphi}{3}+w\right)^3>\left[\frac{16}{27}\varphi^3+\frac{2}{3}w\varphi^2+n^2(\varphi+w)\right]^2$$

时才有 3 个根，整理上式得

$$\frac{1}{3}w^3\varphi^3>\frac{192}{729}\varphi^6+\frac{16}{243}w\varphi^5+\frac{32}{27}n^2\varphi^4+\frac{68}{27}n^2w\varphi$$
$$+\frac{4}{3}w^2n^2\varphi^2+n^4\varphi^2+2wn^4\varphi+n^4w^2. \tag{13}$$

---

① 该法则是说，多项式方程$f(x)=0$的正根的最多个数等于系数变号的次数，而负根的最多个数等于$f(-x)=0$里的系数变号次数。——译者注

如果(13)式成立，则必有

$$\frac{1}{3}w^3\varphi^3>\frac{4}{3}w^2n^2\varphi^2, \tag{14}$$

其中不等号右面的项是(13)式中不等号右边的第五项。条件(14)可化简为

$$w>\frac{4n^2}{\varphi},$$

即

$$w>\frac{4(1+W)^2}{\varepsilon+\eta W}.$$

因此

$$w>4\min\left(\frac{1+W^2}{\varepsilon+\eta W}\right),$$

即

$$w>\frac{8}{\eta}\left(\sqrt{1+\left(\frac{\varepsilon}{\eta}\right)^2}-\frac{\varepsilon}{\eta}\right). \tag{15}$$

这里及以后，我们取值

$$\varepsilon\sim0.06,\quad \eta\sim0.33,\quad V_T\sim8.9408\mathrm{m/s}. \tag{16}$$

其中 $V_T$ 的值由气象站的同僚提供，其它数据由作者测量而得。代入(15)式得

$$w>20.1\cdots,$$

即，为了使淋湿函数在区域$(0,w)$里有最大值和最小值，要求风速在 $i$ 方向的分量超过 178.816m/s。我们排除这种可能性，因为这不现实。

因此淋湿函数可以被假定为在$(0,w)$上单调递减。

## 6. 情形1

这种情形表示条件$w<\varphi$。也就是说，一个人站在雨中，面朝 $i$ 方向，顶部和右侧挨淋的程度比背部严重。

如果$x>w$，则有

$$F(x)=\frac{\varphi+x-w}{x\sqrt{n^2+(x-w)^2}}。\tag{17}$$

类似于前一节的分析得到方程

$$y^3+2\varphi y^2+w\varphi y-n^2(w-\varphi)=0,\tag{18}$$

其中 $y=x-w$。

感兴趣的区域是$y>0$，在应用到情形1的条件下，方程(18)在此区域中没有根。因此在此情形时，如果$x>w$，则$F(x)$无最大值或最小值。

于是，对于所有$x$，$F(x)$是单调递减的，因而，我们再一次得到：最好的策略是跑得尽可能地快。

## 7. 情形2

情形2，即一个人站在雨中，脸朝 $i$ 方向，其背部挨淋的情况比顶部及右侧严重。方程(18)再一次被应用到这种情形，但此时最末一项是负的，从而有根存在。因为当$x\to\infty$时$F(x)\to 0$，所以这和本文开始时提到的考虑是一致的。这种情形时$F(x)$的图形见图1。

存在唯一的点$x=X$使$F(X)=F(w)$。令$X=w+Y$，我们求得

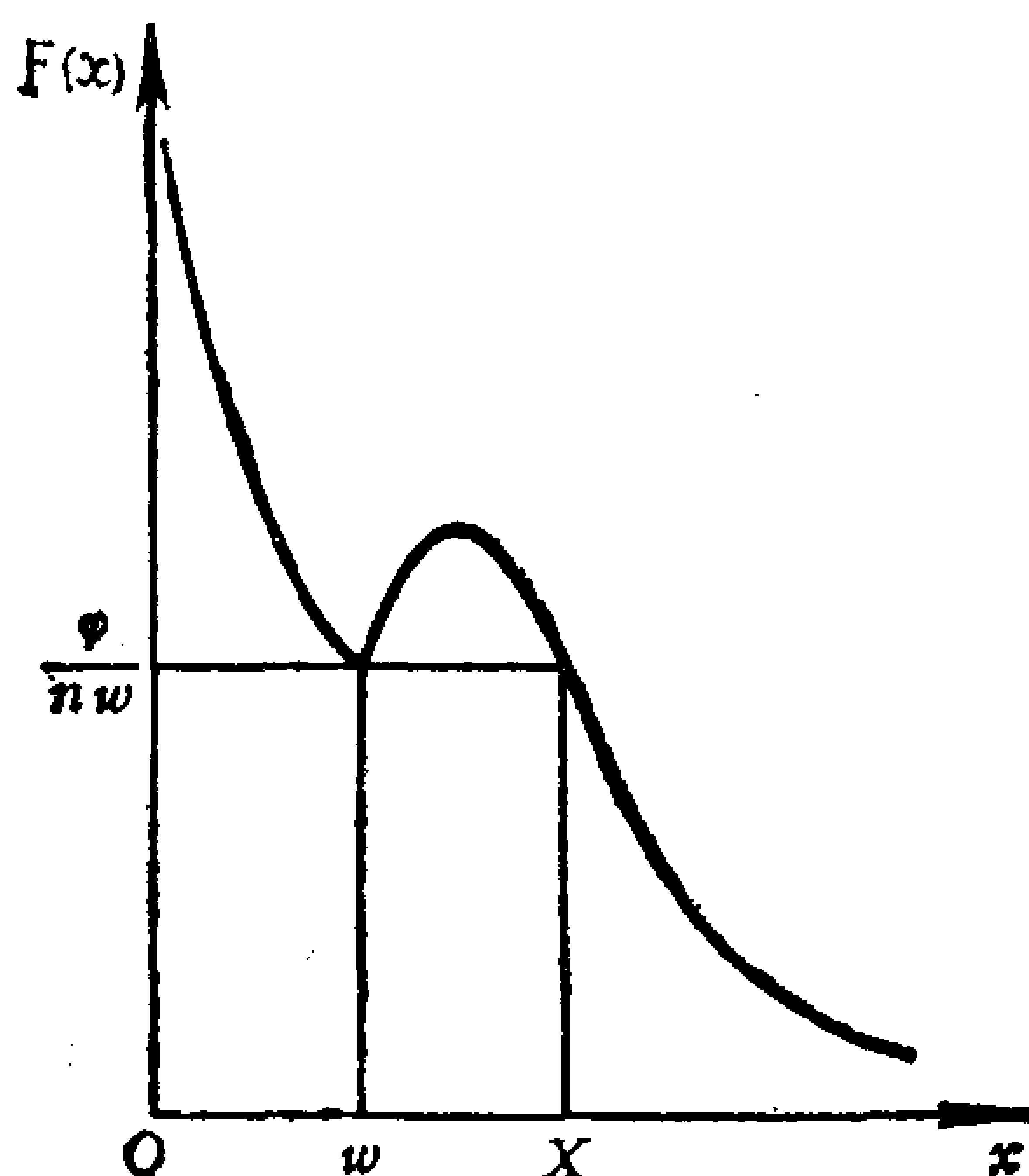



图 1　在情况 2 ($w>\varphi$) 函数 $F(x)$ 的变化

$$\varphi^2Y^3+2w\varphi^2Y^2+(n^2\varphi^2+w^2\varphi^2-w^2n^2)Y-2w\varphi n^2(w-\varphi)=0, \tag{19}$$

由 Discartes 符号法则，它只有一个正根。

在区域$w<x<X$中的速度 $x$ 应该不用，因为如果按这些速度前进，则淋到雨会比该淋的还要多。最佳策略的确定视此人所具有的最快的步行(或者更确切地说是跑)速度而定。如果此速度大于$X$，则他应以他的最快速度奔跑。但是，如果他的最快速度小于$X$，则最好是放慢速度直到仅把头顶及右侧作为挨淋的靶子——即，试图“在雨点之间”行进。

用(16)式中的数据并假定最大可能的行进速度是 8.9408m/s(即 $x=1$)，我们用计算机来研究这种情形。当$W$在 0 和 3 之间取值时，对在$(\varphi,1)$中的一些$w$值，方程(19)可

以被解出。使方程(19)的解等于 $1-w$ 的 $w$ 值是很容易得到的。记此值为$w_c$(即 $w$的临界值——译者注)。因此有可能得到$w_c$和$W$的关系图，这在图 2 中表出。

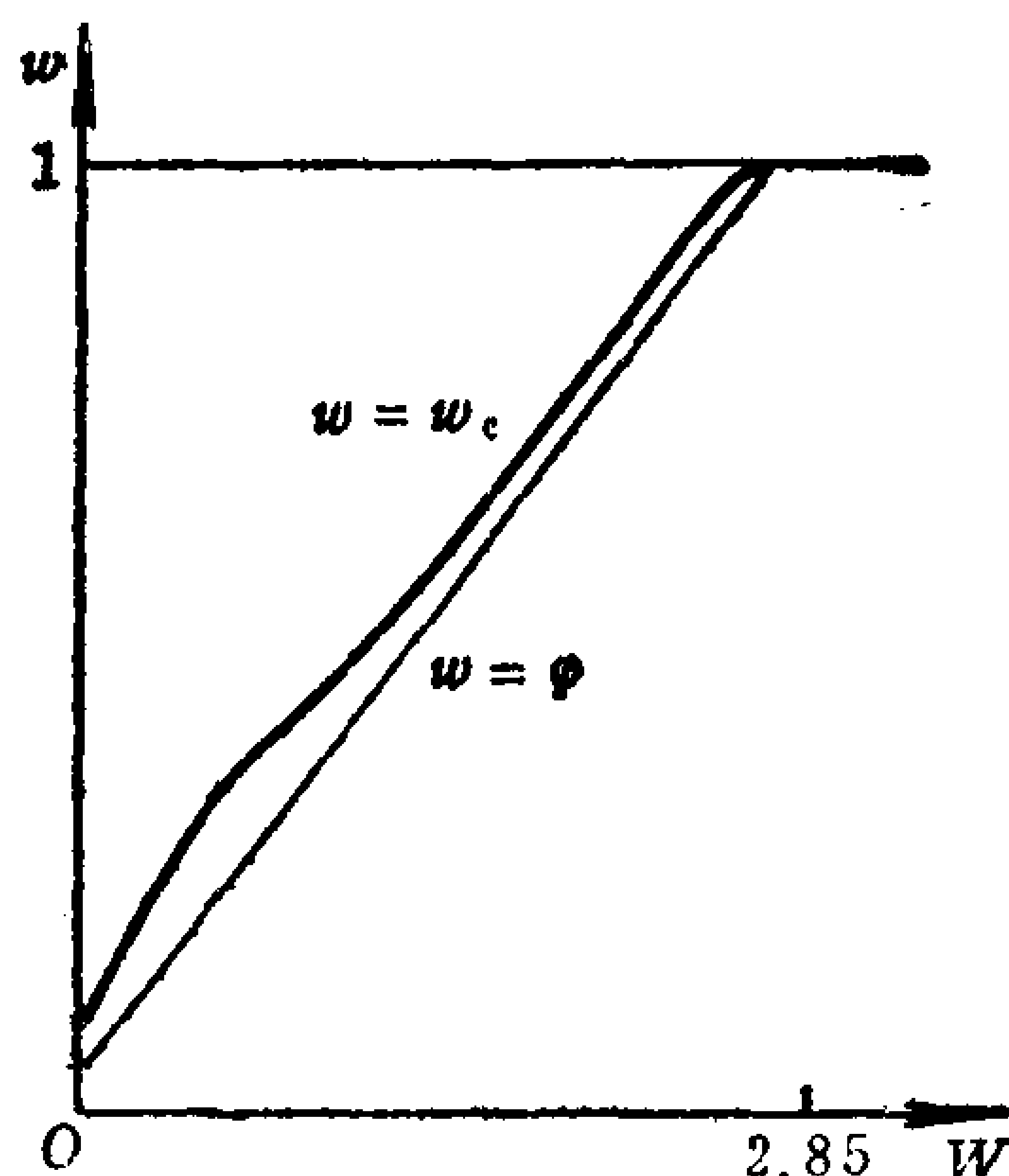



图 2　$w_c$ 对应$W$的依赖性。直线$w=\varphi$用于进行比较

如果 $w>w_c$，则最好的策略是降低行进的速度使和雨速一致，当然，除非$w>1$，这时又是跑得越快越好。

在$\varphi<w<w_c$的情形，令$x=w$得到极小值，如果图省事，这值也可以认为是实际的最小值($x=1$)。

## 8. 灵敏度分析

考虑另外两种可供选择的策略。一种自然的策略是在每一种情形中取 $x$ 等于它的最大值；另一种更高级的策略将在本文中勾划。我们要比较这两种策略的效果。除了情形

$$w_c<w<1$$

以外，这两种策略是相同的，因此这种情形将单独考虑。

如果采取自然的策略，则相应的潮湿函数是$F(1)$，但如果执行高级的策略，由方程(8)得潮湿函数是$\varphi/(nw)$。令

$$R=\frac{nwF(1)}{\varphi}, \tag{20}$$

如果采取自然的策略，$R$量度了遭致的惩罚。例如，如果$R=2$，则一个采用自然策略的人淋湿的程度将2倍于如果采用高级策略的人，等等。

用计算机分析前面研究过的情形.所得结果如图3所示.临界曲线$R=1$必然和图2中的曲线$w=w_c$相同。这提供了计算精确度的有用的检查。由这两种方法产生的曲线检查了涉及到的精确度的极限.(为了更好地展示更感兴趣的临界曲线在图3中仅仅画出了这条曲线的一部分.)

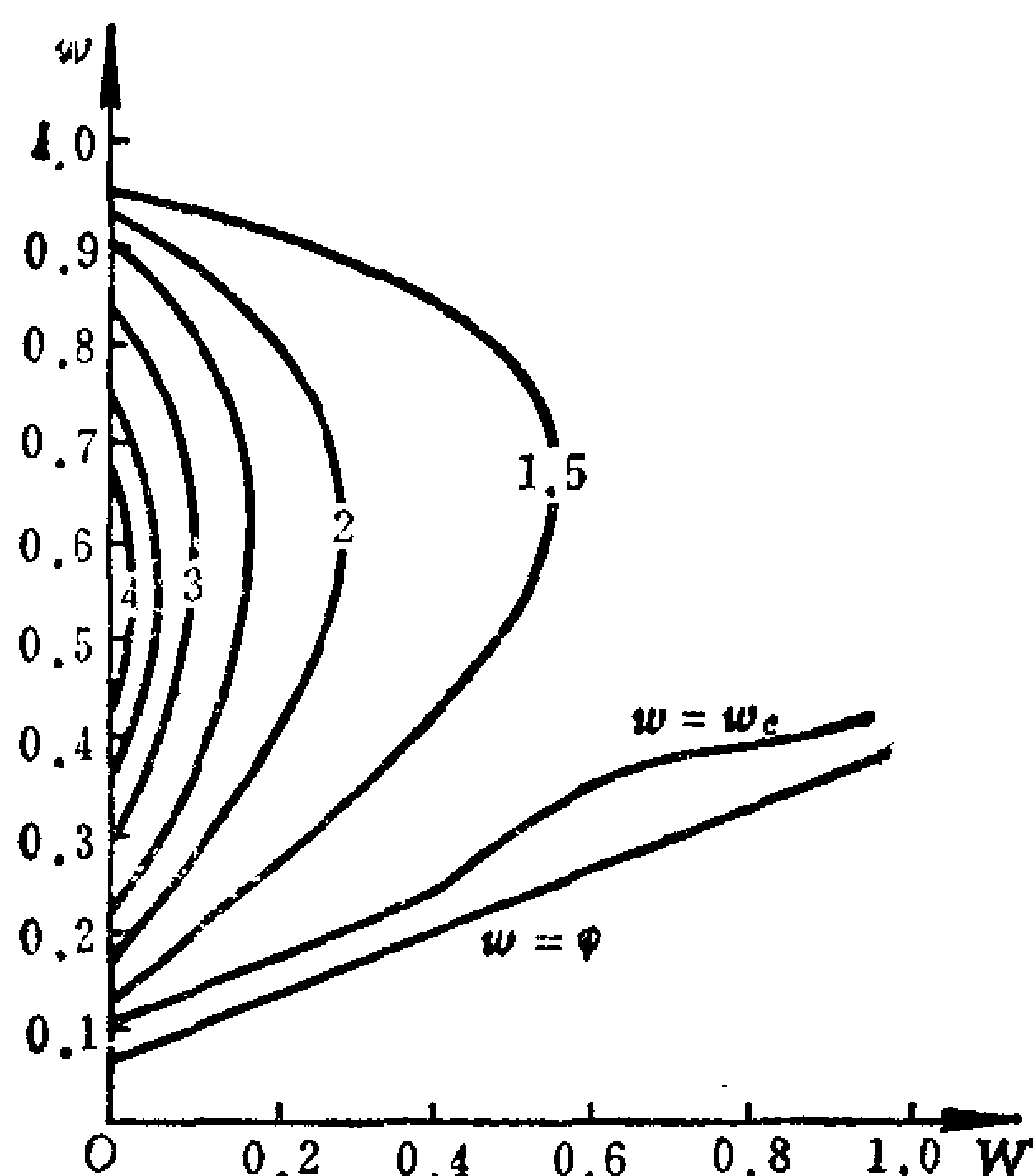



图3 $R$对$w$及$W$的依赖性。临界曲线$R=1$是由曲线$w=w_c$及在$W=0$和$W=2.85$之间的直线段$w=1$组成的

从图中看到，对于小的$W$值，采取自然策略所受到的惩罚是相当严重的，因为这时在 $w$-$W$ 平面上表示出的区域中，$R$ 实际上比 1 大得多。

当$W$增大时，采用高级策略的重要性变小，而且对于较大的$W$值，$R$ 非常接近于 1。例如，当 $W=2$ 时，$R$ 的最大值刚好超过 1.025。这些值相应于有一股相当大的风。但是，也许正因为这样，所以它们与分析问题的关系不大。

要选取的 $x$ 的有价值的值，是那些使淋湿函数取极大值的值（见图 1）.这正是使得 $R$ 取最大值的原因。一个人几乎不可能有意去选择一种走或跑的方法为了得到这个最大值，但他如跑得尽可能地快，兴许会不自觉地达到这个最大值。正是这种情形，对于任意给定的$W$，得到了 $R$ 的最大值。

## 9. 总结

前面，我们分析了一个人在雨中行进的简单模型，也许是最简单实用的模型。对于这个模型，通常的解答是跑得越快越好。采取这种方法，一般可以使被淋湿的部分最少。但是，如果这场雨下得使这个站在雨中脸朝前进方向的人主要是后背受淋，那末存在风速的一个变化范围，对于这些风速，这一方法将无效。这时最好的方法是使人的速度和风速的分量相同。如果风的横向分量不大，则跑得尽可能地快的这种自然的策略产生的结果是：淋湿的程度将 4 倍于采用较高级的方针。当风的这个分量增大时，两种策略之间的差别将逐渐变得不明显。

（朱学贤译，潘承彪校）

# “雨中行”问题的重新考虑①

B.L.Schwartz, M.Deakin

## 引言

在较早的一篇文章里[1]，作者之一研究了下面的问题：一个在雨中走（或跑）的人，应采取何种最佳策略，使挨淋最少。本文改正那篇文章中的技术性错误，同时简化和推广文章中所获的结果。记号及专业术语与[1]中的相同，但为了完整起见，在本文中再叙述一遍所需的定义，使本文成为一篇独立的文章。

## 模型

设 $\boldsymbol{i}$ 是 $\overrightarrow{AB}$ 上的单位向量；$\boldsymbol{k}$ 是方向朝上的单位向量；$\boldsymbol{j}=\boldsymbol{k}\times\boldsymbol{i}$。假定雨是密度均匀的液体，以速度 $V_T$ 落下，并受到速度为 $V_T(w\boldsymbol{i}+W\boldsymbol{j})$ 的水平方向的风吹动。于是，从地面上一个固定点看去，雨的速度是 $V_T(w\boldsymbol{i}+W\boldsymbol{j}-\boldsymbol{k})$。设人的速度是 $\boldsymbol{i}V_Tx$，其中的 $x$ 是要去求的。因此，相对于此人的雨速是 $V_T\{(w-x)\boldsymbol{i}+W\boldsymbol{j}-\boldsymbol{k}\}$。把人设想成一个长方体，六个面中有三面挨淋：前部或背部，右侧或左侧，以及顶部。三部分的面积分别为 $A,\eta A$ 和 $\varepsilon A$。

---

① Walking in the rain, reconsided, *Math. Magazine*, 46(1973), 272—276.

## 淋湿函数

单位时间淋在顶部表面上的雨量正比于面积 $\varepsilon A$ 乘上垂直于该表面的雨速的分量。因此落在这个人头顶上的雨量正比例于 $\varepsilon A$。同样地，袭在侧表面上的雨量正比于 $\eta A|W|$；落在前部或背部的雨量正比例于 $A|w-x|$。最后两个表达式中出现绝对值号，是因为在4个竖直的面中，只有2个面受到雨淋，它们视$W$和 $(x-w)$ 的符号而定。

因此，单位时间里落在一个人身上的总雨量正比于 $\varepsilon+\eta|W|+|w-x|=\varphi+|w-x|$，其中 $\varphi=\varepsilon+\eta|W|$。

当一个人从 $A$ 走到 $B$ 时落在他身上的总雨量和单位时间里落在他身上的雨量成正比，并和他的行进速度成反比——即，和

$$F(x)=\frac{\varphi+|w-x|}{x} \tag{1}$$

成正比，我们称 $F(x)$ 为“淋湿函数”。

这在形式上和上一篇文章中相应的表达式不同。那篇文章中在分母上错误地包含了因子 $\sqrt{(w-x)^2+W^2+1}$。

## 直观的检验

$F(x)$的先前（不正确）的表达式，具有这样的性质：当 $x$ 充分大时它变得任意小。这就意味着，如果一个人能达到足够大的速度，则淋在他身上的雨量能少到他希望的任何水准。显然，这是不对的。因为不管他以多快的速度从 $A$ 走到 $B$，也不可能躲掉这样一些雨量：这些雨构成一个棱柱，其长轴是线段 $AB$，截面等于人的截面。而这个雨量是正常数，

不可能任意小。

而在前面给出的 $F(x)$ 的正确的表达式里，当 $x$ 变得很大时，它接近于一个正的极限值 1，因此它就通过了这一直观的检验。

### 最佳策略

我们来求 $F(x)$ 的最小值。注意到函数 $F(x)$ 的导数在 $x=w$ 处不连续。

$$F'(x)=(\varphi+w)/x^2,\quad x<w \tag{2}$$

及

$$F'(x)=-(\varphi-w)/x^2,\quad x>w. \tag{3}$$

因此，由普通的微积分不能求得 $F(x)$ 的最小值。$F(x)$ 的图形的形状依赖于 $\varphi$ 和 $w$ 的相关的值。如果 $\varphi>w$，则 $F(x)$ 对所有的 $x$ 是单调递减的（见图 1），从而对这个人来说，跑得尽可能地快是他的最佳策略。

另一方面，如果 $\varphi\leqslant w$，则 $F(x)$ 在 $x=w$ 处有最小值（见图 2）。在这种情形时，这个人应以速度 $x=w$ 行进（如果他

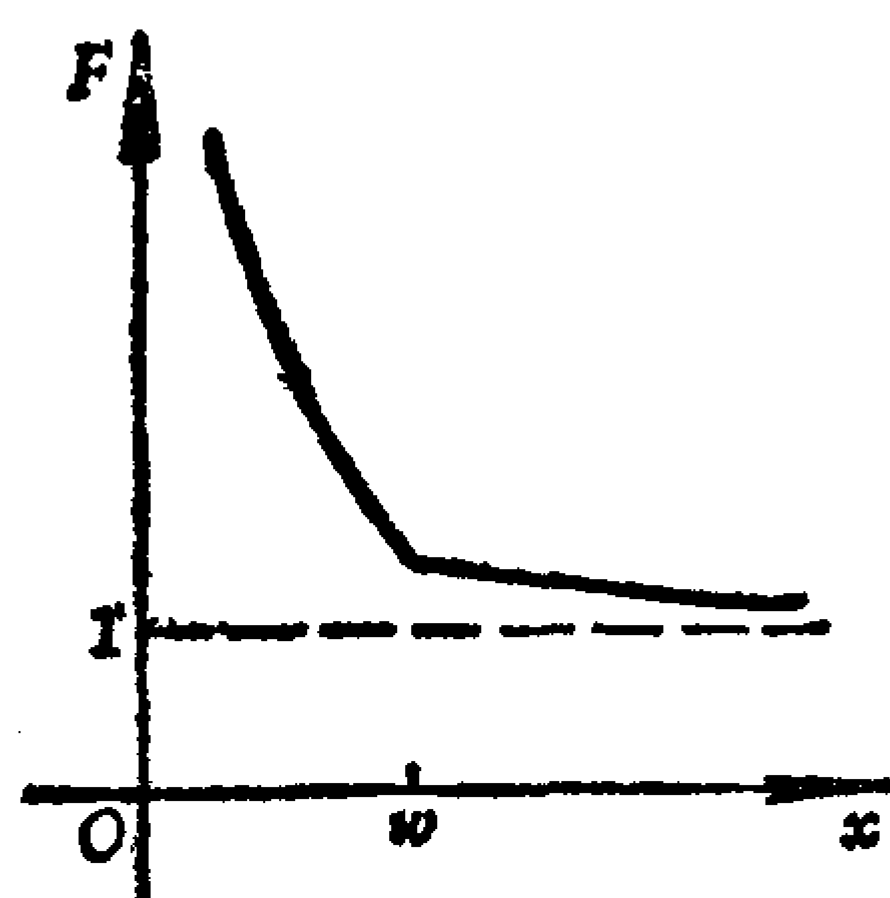



图 1

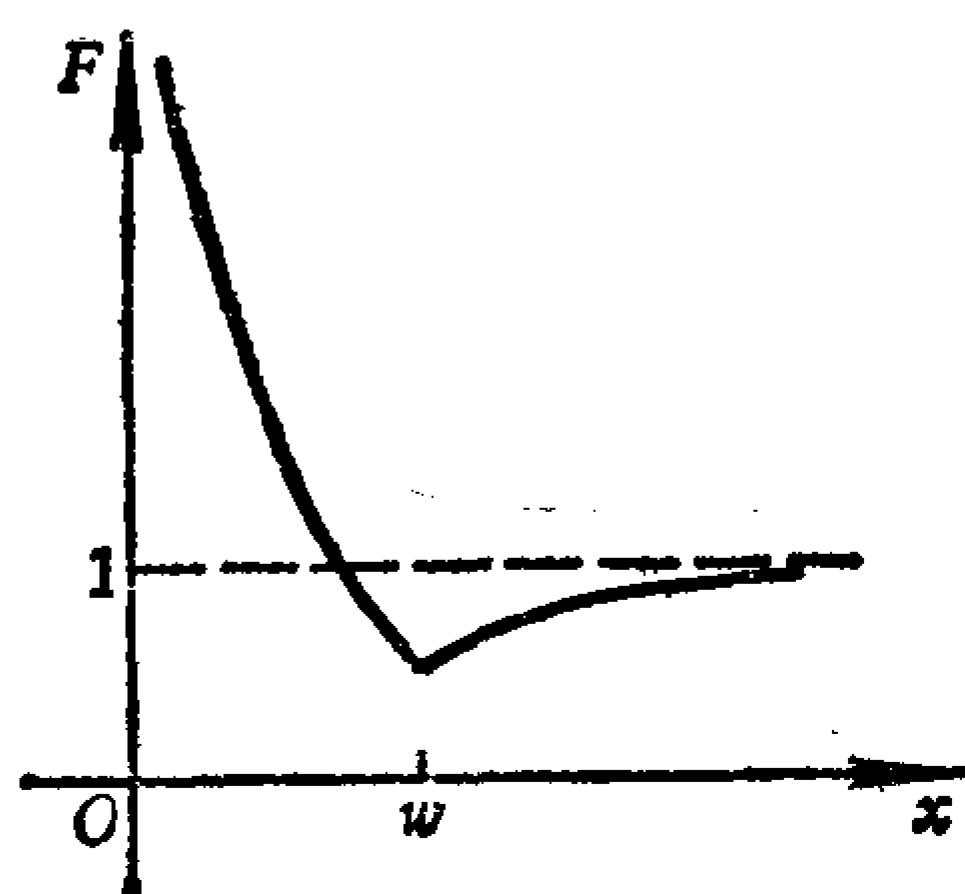



图 2

能做到的话），从而使前部或背部不受雨淋，而仅仅顶部及一侧受淋。如果他的速度不能达到 $w$，即不能达到风的速度，则他的淋湿情形正处于 $F(x)$ 曲线上的单调递减部分，从而再一次地，他应跑得越快越好。

**判定过程**

判定是还是不是以最快速度奔跑要视是否 $\varphi>w$ 而定。如果 $\varphi<w$，即

$$\varepsilon+\eta|W|<w, \tag{4}$$

则最好是使行进速度和风速的有关分量保持一致。不等式(4)定义了 $(w,W)$ 平面上的一块区域，对于它，这一策略是最佳的。在图 3 中，这一策略利用"实"变量 $WV_T$ 和 $wV_T$ 画出，$\varepsilon$，$\eta$ 和 $V_T$ 的值在上一篇文章中给出：$\varepsilon=0.06$，$\eta=0.33$ 及 $V_T=8.9408$ m/s。

将这些值代入不等式(4)得

$$|WV_T|<3wV_T-3.6. \tag{5}$$

判定法则可以简化为：如果速度 $V_T(w,W)$ 的终点落在图 3 的画线区域中，则这个人应以速度 $w$ 行进（如果可能的话）。在所有的其他情形，他都应以最快速度跑动。

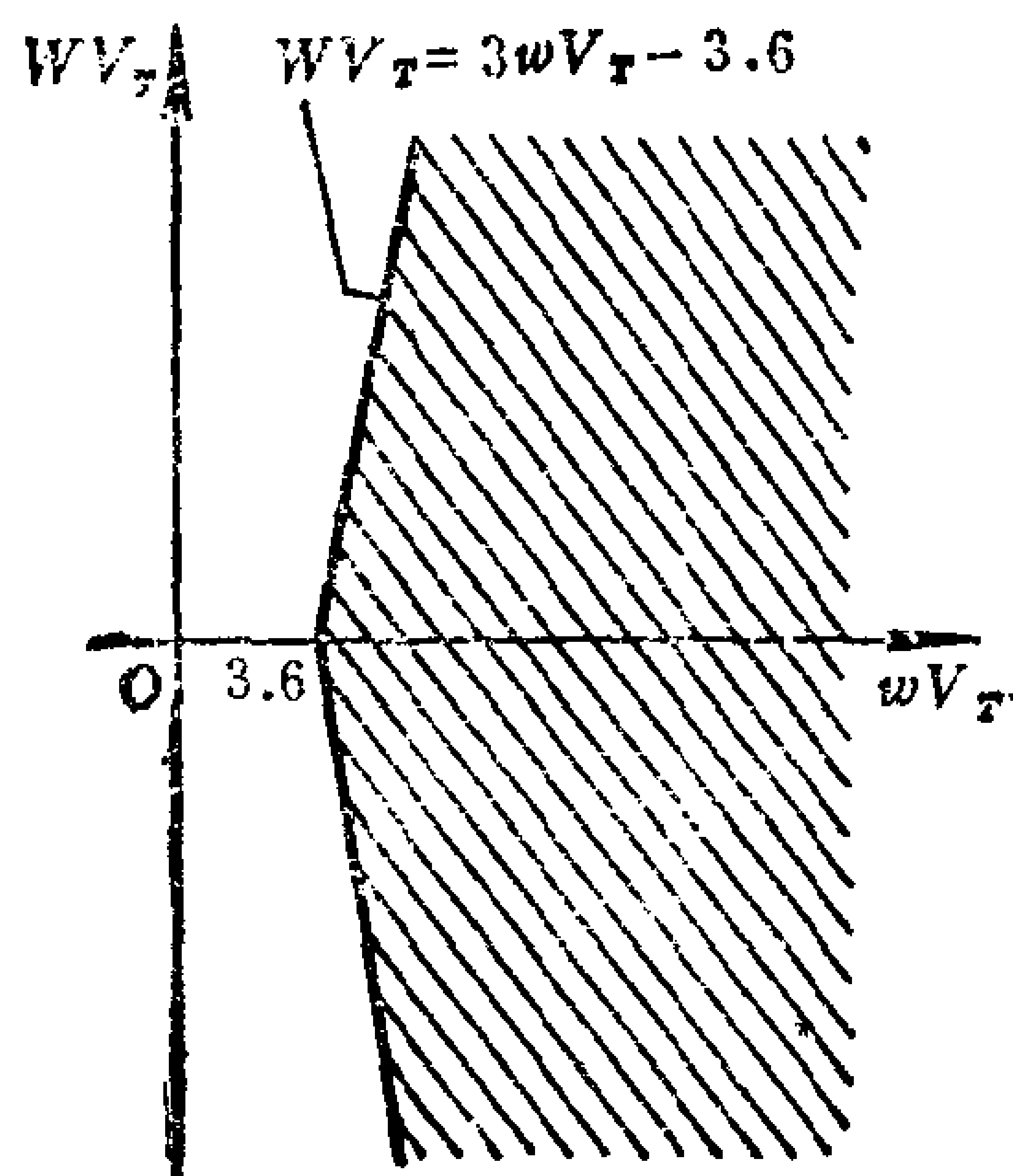



图 3

**灵敏度分析**

设$\varphi<w$，但此人仍以尽可能快的速度奔跑。这人挨淋的情形可以用淋湿函数值$F(X)$来量度，其中$X$是他的最快速度。定义

$$R=\frac{F(X)}{F(w)}. \tag{6}$$

$R$将用于量度如果采用自然策略而不是较高级的策略所受到的惩罚。我们仅仅考虑$X>w$的情形，鉴于在其它情形时两种策略的问题并不出现。

将(1)代入(6)得

$$R=\frac{\varphi+X-w}{X}\cdot\frac{w}{\varphi},$$

由此显然可以发现，当$\varphi$变小时$R$变大。因而最大可能的$R$值是当$\varphi$取最小值$\varepsilon$时发生，即当$W=0$时才出现。在这种境况时

$$R=\frac{w}{X}\left(1+\frac{X-w}{\varepsilon}\right).$$

当

$$X=2w-\varepsilon$$

时$R$取最大值，且最大值为

$$R_{max}=\frac{(X+\varepsilon)^2}{4X\varepsilon}.$$

由在上一篇文章研究过的情形：$\varepsilon=0.06$及$X=1$，得到$R$的最大值为4.68。从而在这完全可能发生的情形下，一个人如

果采用自然策略，即跑得尽可能地快，而另一个人却慢吞吞地走，使其速度与雨速相称，那末前者淋湿的程度是后者的4倍还要多。

## 结束语

我们证明了：纠正上一篇文章[1]的分析中的技术性错误能简化分析但并没有严重地改变结论。修改后的分析提供了求 $|x|$ 型函数的最小值的一个有理由接受的且似乎符合实际的讨论，这对于初等微积分的教师们可能会有一些用处。本文中内含的简化猜测了更符合实际的模型的可行性；例如，也许可以将雨都恰好在同一个方向上落下来，或者都有同样的速度这样一些假设减弱。然而，我们还没有作这种推广的任何企图。

## 参 考 文 献

[1] M.A.B.Deakin, Walking in the rain, *Math. Magazine*, 45 (1972), 246—253.

（朱学贤译，潘承彪校）

# 邮 票 问 题①

Ronald Alter, Jeffrey A. Barnett

假如规定每个信封上至多只能贴 $h$ 张邮票，而你拥有 $k$ 种不同的整数面值的邮票，试就给定的 $h$ 及 $k$ 确定最大的整数 $n=n(h,k)$ 使得你手头上的邮票能够分别组成值为 $1,2,\cdots,n$ 的各种邮资（每次组成不得超过 $h$ 张邮票）。另外，求出满足这个条件的所有由 $k$ 种整数面值构成的集合（称之为解集合）。

通常地，我们在上述的解集合中加入面值为 0 的邮票。这样，邮票问题中的条件便可改为要求每个信封恰需贴上 $h$ 张邮票。例如，当 $h=2$ 及 $k=3$，则 $n=n(2,3)=8$，且相应的唯一解集合为 $\{0,1,3,4\}$（解集合中的整数表示邮票的面值）。面值分别为 $1,2,\cdots,8$ 的邮资可如下构成：

$$1=0+1,\quad 2=1+1,\quad 3=0+3,\quad 4=0+4,$$
$$5=1+4,\quad 6=3+3,\quad 7=3+4,\quad 8=4+4.$$

解集合未必都唯一，可以有许多。比如，$n=n(2,6)=20$，它的解集合有 5 个：$\{0,1,2,5,8,9,10\}$，$\{0,1,3,4,8,9,11\}$，$\{0,1,3,4,9,11,16\}$，$\{0,1,3,5,6,13,14\}$，及 $\{0,1,3,5,7,9,10\}$。

---

① A Postage Stamp Problem, *Amer. Math. Monthly*, 87(1980), 206—210.

显然，$n(1,k)=k$ 且相应解集合为$\{0,1,\cdots,k\}$；$n(h,1)=h$ 且相应解集合为$\{0,1\}$。Stöhr[44]，Henrici[12]，及Stanton等人[43]独立地给出

$$n(h,2)=\lfloor(h^2+6h+1)/4\rfloor,$$

其中$\lfloor x\rfloor$表示不超过$x$的最大整数。如果$h$为奇数，则它的唯一的解集合是$\left\{0,1,\frac{1}{2}(h+3)\right\}$；如果$h$为偶数，则存在两个解集合：$\left\{0,1,\frac{1}{2}(h+2)\right\}$及$\left\{0,1,\frac{1}{2}(h+4)\right\}$。

其它情况中仅当$k=3$时有一个几乎使问题解决的解(即$n(h,k)$的上下界很接近)。Hofmeister[17]证明了

$$\frac{4}{81}h^3+\frac{2}{3}h^2+\frac{66}{27}h\leqslant n(h,3)\leqslant\frac{4}{81}h^3+\frac{2}{3}h^2+\frac{71}{27}h-\frac{1}{81}\quad(h\geqslant 34),$$

其下界在$h\equiv 0\pmod 9$时达到。Klotz[20,21]及Henrici[12]也分别提到过相似的、但较弱的下界。

1936年，Rohrbach[37]对固定的$h$及足够大的$k$得到$n(h,k)$的一个渐近界

$$(k/h)^h\leqslant n(h,k)\leqslant k^h/h!+O(k^{h-1})①.$$

其下界的证明是构造性的，而上界的证明则是平凡的。因为$(k+1)$个不同元素所组成的重数不受限制的多重集的全部$h$-

① 设函数$f(x,y,\cdots)\geqslant 0$，符号$O(f(x,y,\cdots))$表示一个函数$g(x,y,\cdots)$，它满足条件：存在一个正常数$C$，使得$|g(x,y,\cdots)|\leqslant Cf(x,y,\cdots)$。

组合数是$\binom{k+h}{h}$，因此，$n(h,k) \leqslant \binom{k+h}{h}-1$。这里所以要减1是由于要排除所取$h$张邮票的面值都是零的情形。

Hofmeister（见文献[19,P.112]）应用R.Widecker的一个未发表的结果

$$n(h,k) \geqslant (4/3)^{\lfloor h/3 \rfloor}(8/7)^{\lfloor (h-3\lfloor h/3 \rfloor)/2 \rfloor}(k/h)^h - O(k^{h-1}),$$

同样导出了上述著名的下界.Hofmeister（见文献[19,P.104]）还对固定的$k \geqslant 3$及足够大的$h$给出了相应的上下界：

$$2^{\lfloor k/4 \rfloor}(4/3)^{\lfloor (k-4\lfloor k/4 \rfloor)/3 \rfloor}(h/k)^k + O(h^{k-1})$$
$$\leqslant n(h,k) \leqslant h^k/k! + O(h^{k-1}).$$

对$h=2$，Rohrbach[37]发表了一个非凡的上界

$$n(2,k) \leqslant \frac{1}{2}(1-0.0016)k^2 + O(k).$$

此后，这一结果得到Klotz[20,21,22]的改进。他将0.0016推进到0.0369。

Moser[27]，Riddell[36]，Salié[38]，及Moser和Riddell[28]也分别做了另外一些有关$n(h,k)$的上界的工作。对于足够大的$k$的情形，至今为止最好的上界要归功于Moser等人（见文献[29]），他们发现了

$$n(h,k) < (1-b_h)\frac{k^h}{h!},$$

其中$b_3=0.0221$，及$b_4=0.0115$。此外，当$h \geqslant 5$时$b_h=(1.02f(h))^h$，当$h \geqslant 8$时$b_h=(1.1f(h))^h$，这里$f(h)=\cos(\pi/h)/(2+\cos(\pi/h))$。

Richard K.Guy曾提出，对于充分大的$h$，$n(h,k)$可由有限个关于$h$的$k$次多项式来表示。例如，Stöhr所给的$k=2$

的解可以表成

$$n(h,2)=(h^2+(3+3c)h+d)/4,$$

其中 $c=d\equiv h \pmod 2$。Guy 对 $k=3$ 及 $h\geqslant 20$ 的猜想是

$$n(h,3)=(4h^3+54h^2+(204+3C_r)h+d_r)/81,$$

其中 $h\equiv r \pmod 9$，且 $C_r$ 及 $d_r$ 由下表给出：

$$r=-4,-3,-2,-1,0,1,2,3,4,$$
$$C_r=0,1,3,0,-2,0,3,1,0,$$
$$d_r=46,-81,-1,-170,0,62,-26,0,-154.$$

邮票问题由来已久。然而，我们所能找到的有关这问题的最早文献是 Rohrbach [37]。一些邮票问题的特殊形式曾出现在一些趣味数学及其它科普书刊上。请参见 Sprague [42,问题18]，Gardner[3,问题 4]，及 Legard [24]。

与解决 $n(2,k)$ 有密切关系的一个问题是通过解集合的成员之差来表示整数 $1,2,\cdots,n$。Miller [26] 及 Leech [23] 描述了这一问题。

Alter 及 Barnett [1] 叙述了 $n(2,k)$ 问题在计算机的变质寄存器的最优分配上的一个应用。Hargraves [6] 又给出了另一个应用。为了设计可联贮存器的最佳接线图，他利用了关于 $n(h,k)$ 的解集合。

至今，所有发表过的求解 $n(h,k)$ 的算法都是关于 $h$ 及 $k$ 的指数型算法。需要在计算机上花费数千小时，以得到 $n(h,k)$ 的值。表 1 列出了除能由简单表达式给出的值（亦即 $h=1$ 或 $k=1,2$）之外的所有已知的 $n(h,k)$ 的值。这里所给出的值最初是发表在 Stohr [44]， Henrici [12]， Lunnon [25]，

Seldon[39,40],Phillips[35]，以及 Alter 和 Barnett[1] 上的。后来，经过核实，这些结果再由 Stanton 等人[43]及 Heimer 和 Langenbach[11] 给出，另外，Henrici 曾发表了当 $k=14$，…，18 时 $n(2,k)$ 的值分别为 80，92，104，116，及 128. 他获得这些值所用的工具是一种未加证明的修枝式的经验方法。因此，在更可信的方法被采用之前，这些值应该被视为是一个下界。对 $k\leqslant 47$，$n(3,k)$ 的值是由 John A. Bate 计算的。十分感谢他允许我们在此发表他的这些数据。

表1 已知的 $n(h,k)$ 的值

| h \ k | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 8 | 12 | 16 | 20 | 26 | 32 | 40 | 46 | 54 | 64 | 72 |
| 3 | 15 | 24 | 36 | 52 | 70 | 93 | 121 | 154 | | | |
| 4 | 26 | 44 | 70 | 108 | 162 | 220 | | | | | |
| 5 | 35 | 71 | 126 | 211 | | | | | | | |
| 6 | 52 | 114 | 216 | 388 | | | | | | | |
| 7 | 69 | 165 | 345 | | | | | | | | |
| 8 | 89 | 234 | 512 | | | | | | | | |
| 9 | 112 | 326 | 797 | | | | | | | | |
| 10 | 146 | 427 | | | | | | | | | |
| 11 | 172 | 547 | | | | | | | | | |
| 12 | 212 | 708 | | | | | | | | | |
| 13 | 259 | 873 | | | | | | | | | |
| 14 | 302 | 1094 | | | | | | | | | |

九种特殊情况的研究值得注意。Wegner 和 Doig[45] 探索了对称的票面值集合。设 $\gamma=\{a_0=0<a_1<\cdots<a_k\}$ 是一个票面值之集。我们称 $\gamma$ 是对称的如果 $\gamma$ 中相邻的元素之差所

表 2

| $k$ | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $n(3,k)$ | 354 | 418 | 476 | 548 | 633 | 714 | 805 | 902 | 1012 | 1127 | 1254 |
| $k$ | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 |
| $n(3,k)$ | 1382 | 1524 | 1678 | 1841 | 2010 | 2188 | 2382 | 2584 | 2801 | 3020 | 3256 |
| $k$ | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 |
| $n(3,k)$ | 3508 | 3772 | 4043 | 4326 | 4628 | 4941 | 5272 | 5606 | 5960 | 6334 | 6723 |

构成的序列具有回文性(即顺读与倒读一样)。我们已经知道，除$k=10$之外，均存在$n(2,k)$的对称解集合。Rohrbach(见文献[37])研究了对称集合的一个限制类，且由此导出了他的渐近界。

Henrici[12]去掉所有$a_i \geqslant 0$的限制。他找到$n(2,7)$的一个解集合$\{-1,2,3,4,10,11,12,15\}$，并声明$n(2,7)$的值为27。而表1给出的值却是$n(2,7)=26$。注意，该声明是正当的。因为正规问题的提法是允许有一个不算数的0元素，所以正规的结果是0,1,…,26，而该声明的结果是1,2,…,27。并且在新的条件下，Henrici对$n(2,10)$找到一个对称(且是唯一)的解集合$\{-1,1,2,4,8,12,16,20,22,23,25\}$，能组成的票面值是从0到48。这个值域不含$-1$是由于每个票面值均需由解集合中的两个元素来构成。

Alter及Barnett[1]对$h=k$导出了一个很有意义的下界，即$n(h,h)\geqslant f_{2h}-1$，其中$f_i$是第$i$个Fibonacci数。

自从Rohrbach首次提出邮票问题以来，我们朝着求解方向已经取得了巨大的进展。尽管如此，仍有许多关键问题

尚待解决。

**问题1** $n(h,k)$的上下界是否可以改进？我们所已知的最好的上下界之间的差距较大，未尽人意。显然，只要没有找到$n(h,k)$的简单表达式，总会有改进上下界的余地。

**问题2** 在$n(h,k)$与$n(k,h)$之间是否存在简单的联系？

**问题3** 作为$h$和$k$的函数$n(h,k)$，相应的解集合之重数(即不同解集合的数目)是什么？

**问题4** 设$\nu=\{a_1,\cdots,a_k\}$，且定义$n(h,\nu)$为最大的整数$n$使得所有整数$1,2,\cdots,n$都能够表成不超过$h$个的$a_i$之和。那么，$n(h,\nu)$可否有一个简单的表达式？注意

$$n(h,k)=\max_{\nu\in U_k} n(h,\nu),$$

其中$U_k$是由所有$k$种不同面值之邮票集构成的族。

对$n(h,\nu)$的认识，必然对改进$n(h,k)$的估计会有巨大的帮助。我们所已知的下界便是通过对$U_k$的限制使得$n(h,\nu)$很容易表示而获得的。

**问题5** 设$\{a_1,\cdots,a_k\}$是$n(h,k)$的一个解集合。那么，$a_i$的上下界是关于$h$及$k$的一个什么样的函数？相对于$a_i$，$a_{i+1}$的数量级是什么？

**问题6** 假如允许有负票面值和分数票面值，则$n(h,k)$将会是什么样的呢？

**问题7** 当$h$和$k$为何值才存在对称解？

**问题8** 是否存在求$n(h,k)$及相应之解集合的多项式-时间的算法？

参考文献目录中包含了几篇文中未提到的文章。由于邮票问题的提法有好几种，且有关问题的名目繁多，这就使得

我们搜集文献的工作变得异常的困难。早在1955年，Stöhr（见文献[44]）就曾做过总结性的工作。

M.L.V.Pitteway 及 R.K.Guy 帮助我们搜集了许多文中提到的文章，N.Goldman 也帮助我们翻译了一些文章。作者特此一并致谢。

## 参 考 文 献

[1] R.Alter and J.Barnett, Remarks on the postage stamp problem with applications to computers, Congressus Numerantium 19, Proc. Eighth Southeastern Conf. on Combinatorics, Graph Theory, and Comput., Baton Rouge, La., 1977, pp.43—59.

[2] M.Djawadi, Kennzeichnung von Mengen mit einer additiven Minimaleigenschaft, Diss. Joh. Guttenberg-Univ., Mainz, 1974.

[3] M.Gardner, Mathematical games, A collection of short problems and more talk of prime numbers, Scientific American, 210 #6(June 1964), Problem 4, p.116; 211 #7 (July 1964), p.114.

[4] N. Hämmerer, Reichweite von Extremalbasen bei fester Ordnung, Diss.Joh.Guttenberg-Univ., Mainz, 1974.

[5] N.Hämmerer and G. Hofmeister, Zu einer Vermutung von Rohrbach, J. Reine Angew. Math., 286/287(1976), 239—247.

[6] R.F. Hargraves, Jr., Application of the postage stamp problem to associative cache memory design, Dartmouth College, Hanover. N.H., 1971.

[7] E. Härtter, Basen für Gitterpunktmengen, J. Reine Angew.Math., 202(1959), 153—170.

[8] ——, Einige Abschätzungen für Abschnittsbasen, J. Reine Angew. Math., 205(1960/61), 82—90.

[9] ——, Eine Bemerkung über Basen, Math. Ann., 165(1966), 24—25.

[10] ——, Additive Zahlentheorie, Vorlesung anider, Joh. Guttenberg-Univ., Mainz, Sommersemester 1973.

[11] R.L.Heimer and H. Langenbach, The stamp problem, J.Rec.Math., 7(1974), 235—250.

[12] A.Henrici, The coins problem, Part 1, Diss., Diploma in Num.Anal and Auto.Comput, Corpus Christi College, Cambridge, 1965.

[13] G. Hofmeister, Methoden zur Abschätzung von $s(h,k)$ und $g(h,k)$ nach unten, Diplomarbeit, Freie Univ. Berlin, 1963.

[14] ——, Über eine Menge von Abschnittsbasen, J. Reine Angew.Math., 213(1963), 43—57.

[15] ——, Zu einem Problem von Frobenius, Det Kgl. Norske Vidensk. Selsk. Skr.(1966) Nr.5, 1—37.

[16] ——, Über eine Menge von Abschnittsbasen 2, Det Kgl. Norske Vidensk. Selsk. Forhandlinger, 39(1966), 60—65.

[17] ——, Asymptotische Abschätzungen für dreielementige Extremalbasen in natürlichen Zahlen, J. Reine Angew. Math., 232(1968), 77—101.

[18] G.Hofmeister and H. Schell, Reichweiten von Mengen natürlicher Zahlen I, Det Kgl. Norske Vidensk. Selsk. Skr.(1970) Nr. 10, 1—5.

[19] G.Hofmeister, Endliche additive Zahlentheorie, Kapitel I, Das Reichweitenproblem, Joh. Guttenberg-Univ., Mainz, 1976.

[20] W.Klotz, Extremalbasen mit fester Elementeanzahl, Diss. Tech. Univ. Carolo-Wilhelmina, Braunschweig,

1968.

[21] ——,Extremalbasen mit fester Elementeanzahl,J. Reine Angew. Math., 237(1969), 194—220.

[22] ——,Eine obere Schranke für die Reichweite einer Extremalbasis zweiter Ordnung, J. Reine Angew.Math., 238(1969),161—168.

[23] J.Leech, On the representation of $1,2,\cdots,n$ by differences, J.London Math. Soc., 31(1956),160—169.

[24] A.Legard, Brain-Teaser, Sunday Times, 23 Dec.1962, and 20 Jan. 1963.

[25] W.F.Lunnon, A postage stamp problem, Comput. J., 12(1969),377—380.

[26] J.C.P.Miller, Difference bases; Three problems in additive number theory, in Symposium on Computers in Number Theory,Atkin and Birch,eds.,Academic Press, 1971, pp.299—322.

[27] L.Moser, On the representation of $1,2,\cdots,n$ by sums, Acta Arith., 6(1960),11—13,

[28] L.Moser and J. Riddell, On additive $h$-bases for $n$, Colloq.Math., 9(1962),287—290.

[29] L.Moser, J.R.Pounder, and J.Riddell, On the cardinality of $h$-bases for $n$, J.London Math. Soc., 44(1969), 397—407.

[30] A.Mrose, Die Bestimmung der extremalen regularen Abschnittsbasen mit der Hilfe einer Klasse von Kettenbruchdeterminanten, diss.Freie Univ., Berlin, 1969.

[31] ——, Eine untere Schranke für die Reichweite von Extremalbasen dritter Ordnung.J.Reine Angew. Math., 261(1973),216—220.

[32] ——, Ein rekursives Konstruktionsverfahren für Abschnittsbasen, J. Reine Angew. Math., 271(1974),214—

217.

[33] ——, Untere Schranken für Extremalbasen fester Ordnung, I, J. Reine Angew. Math. (im Druck).

[34] H.H. Ostmann, Additive Zahlentheorie I, Berlin-Göttingen-Heidelberg, Springer, 1956,

[35] B.P. Phillips, The postage stamp problem, Comput. J., 19(1976), 93.

[36] J. Riddell, On bases for sets of integers, thesis, Univ. of Alberta, 1960.

[37] H. Rohrbach, Ein Beitrag zur additiven Zahlentheorie, Math. Z., 42(1936), 1—30.

[38] H. Salie, Reichweite von Mengen aus drei natürlichen Zahlen, Math. Ann., 165(1966), 196—203,

[39] J.L. Seldon, The postage stamp problem, Comput. J., 15(1972), 361.

[40] ——, The ten stamp problem, Brunel Univ., Middlesex, 1973.

[41] J. Smith, A note on the postage stamp problem (talk only), Combin. Math. III, Proc. Third Australian Conf., Brisbane, 1974.

[42] R.P. Sprague, Recreations in Mathematics (transl. by T. H. O'Beirne), Dover, New York, 1973.

[43] R.G. Stanton, J.A. Bate, and R.C. Mullin, Some tables for the postage stamp problem, Proc. Fourth Manitoba Conf. on Numerical Math., 1974, pp. 351—356.

[44] A Stöhr, Gelöste und ungelöste Fragen über Basen der natürlichen Zahlenreihe I und II, J. Reine Angew. Math., 194(1955), 40—65, 111—140.

[45] P. Wegner and A. Doig, Symmetric solutions of the postage stamp problem, Rev. Franc. Recherche Op., 41 (1966), 353—374.

[46] R.Windecker, Eine Abschnittsbasis dritter Ordnung, Det Kgl. Norske Vidensk.Selsk. Skr.(1976)Nr.9, 1—3.
[47] J.Zöllner, Über Mengen natürlicher Zahlen, für die jede euklidische Darstellung eine minimale Koeffizientensumme besitzt, Diplomarbeit, Joh.Guttenberg-Univ., Mainz, 1974.

(陈赐平译，方祖耀校)

# 初等数学问题的魅力①

P.R.Scott

本文将介绍一些数学观念，它们是初等的，容易为业余爱好者所接受，但是要彻底把它们搞清楚，却需要有广泛而深厚的数学知识。

首先我们观察一块方形的木板，上面钉有布成正方形栅格的钉子(至少在澳大利亚这种装置被用作几何的示教板)。绕着钉子绷上一根橡皮筋就能构造出各种各样的多边形(见图1)。有两个基本概念需要说明：平面上的整数格是指有整数坐标的点的集合，格多边形是每一个顶点为格点的多边形。

现在要问：这些格多边形有什么性质？关于格多边形发现了什么好的结果？这些结果可以作哪些推广？

## 1. Pick 定理

或许最著名的结果是G.Pick在1900年所发现的定理。

**定理** 设$P$是一个简单的(即它的边不自交)格多边形，在边上有$B$个格点，在内部有$I$个格点，则$P$的面积

$$A(P)=\frac{1}{2}B+I-1.$$

① The fascination of the elementary, *Amer. Math. Monthly*, 94 (1987), 759—768.

例如在图 1 中，$B=10$，$I=1$，而 $A(P)=5$。

在 Coxeter 的书（见文献[3]）中有这个定理的证明，它分成两部分：

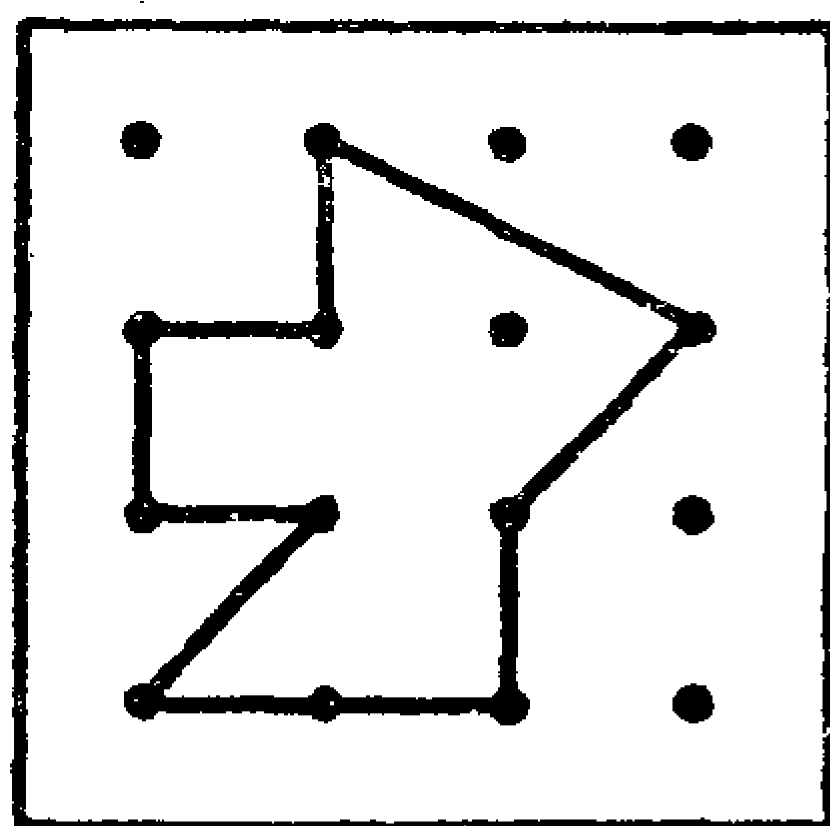


图 1

（1）如果$I=0$，并且 $B=3$，则这是一个内部没有格点的格三角形，以后称之为初等三角形。显然，对于初等三角形$T$，Pick 定理是正确的，此时 $A(T)=\frac{1}{2}$。

（2）上面的面积公式是可加的。它的意思是指：如果 Pick 定理对于格多边形 $P_1, P_2$ 成立，而且 $P_1, P_2$ 有一条公共边界，则 Pick 定理对于去掉公共边界所得的格多边形 $P_1+P_2$ 也是成立的（图 3）。

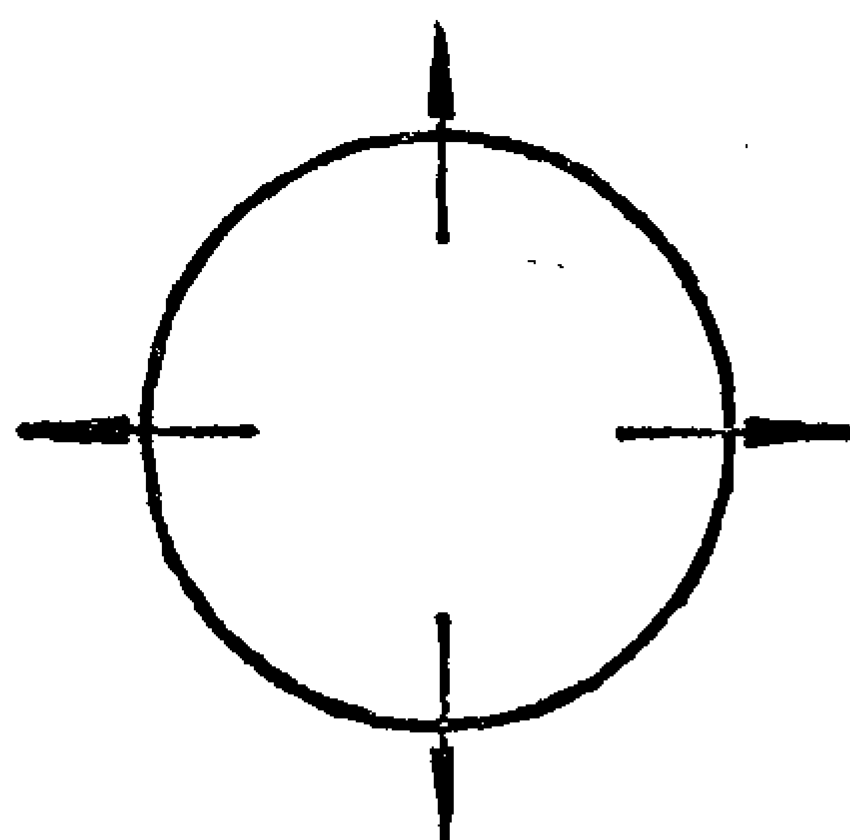



图 2

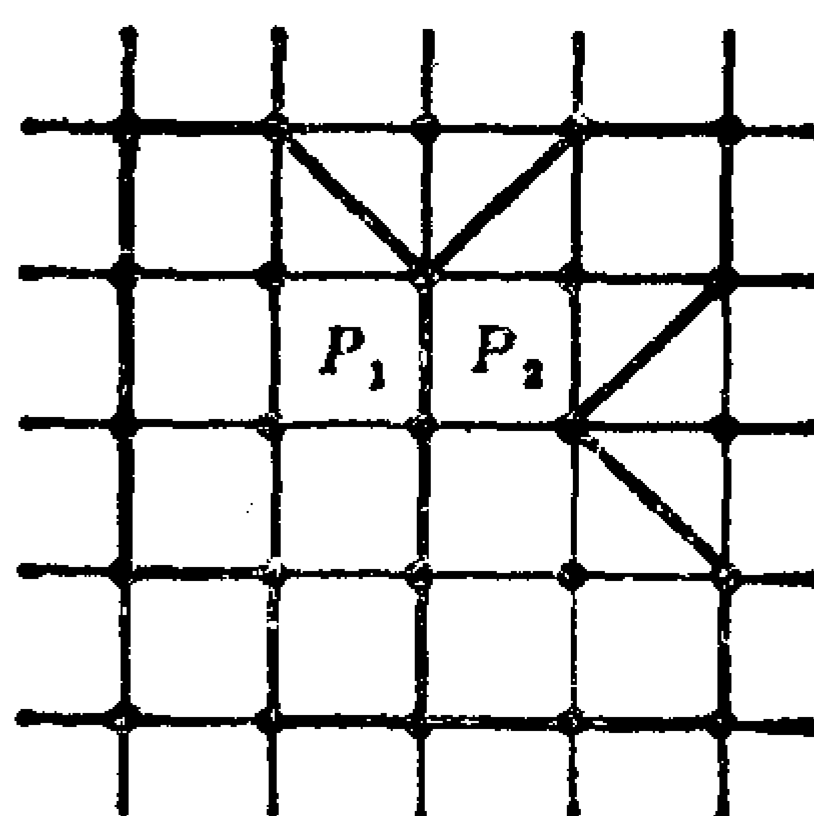



图 3

因为每个格多边形都能由初等三角形构成，因此通过归纳法利用上面的(1)，(2)便可证明 Pick 定理。

如果我们知道了每个初等三角形的面积为$\frac{1}{2}$，那么Pick定理等价于证明：在格多边形$P$中，包含的初等三角形的个数

$$N(P)=B+2I-2.$$

事实上到这一步格就成为多余的了。因为对于有$B$个边界点，$I$个内点的一般的三角形剖分，$N(P)$恰好是把$P$分成互不重叠的三角形的个数（见图4）。

进而可以证明关于$N(P)$的剖分公式等价于著名的Euler公式

$$V-E+F=2,$$

其中$V,E,F$分别是一个平面图形的顶点数、边数和面数。

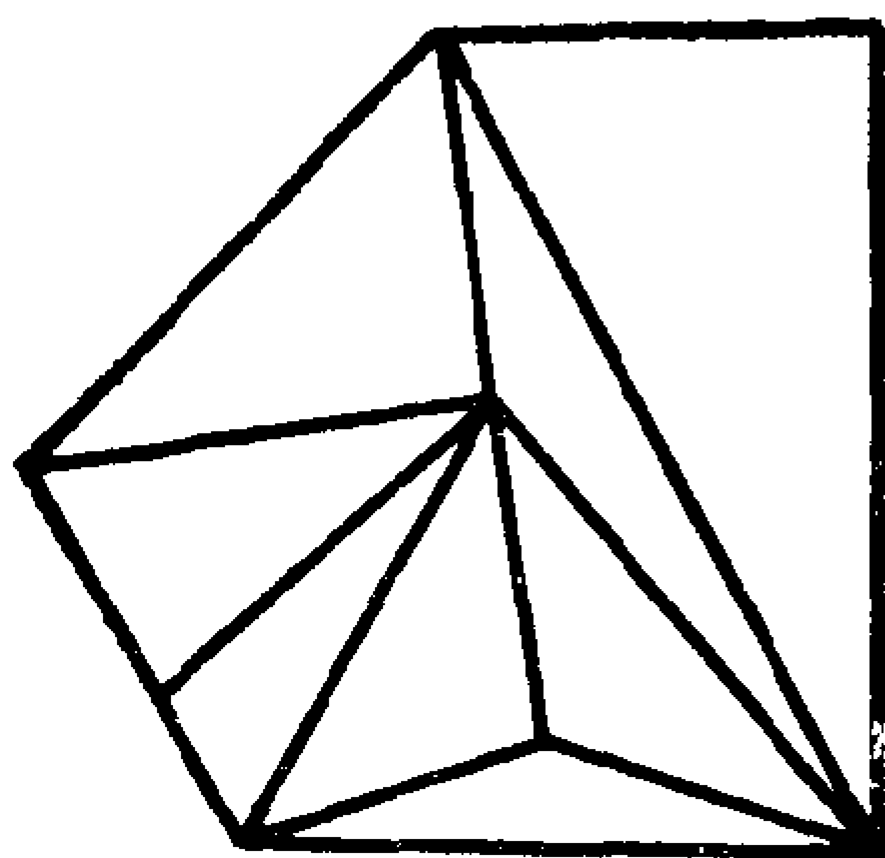


图4　$B=6$，$I=2$，$N(P)=8$

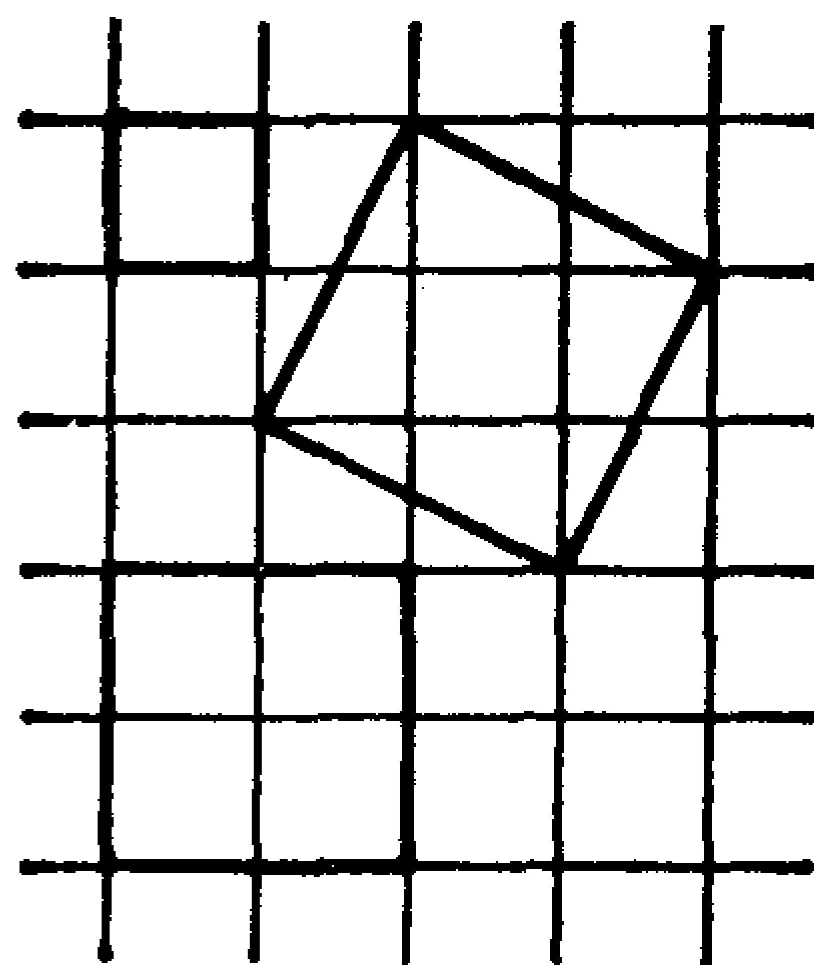


图　5

探求这些关系的有一大批人，其中包括Funkenbusch[6]，Gaskell，Klamkin和Watson[7]，Haigh[10]，Honsberger[12]，Liu[16]，Niven和Zuckerman[19]，Varberg[29]，

以及 A.M.Yaglom 和 I.M.Yaglom [33]。

## 2. 正多边形

各边都相等，且各顶角都相等的多边形称为正多边形。

正方形显然是一个正多边形，同时是一个格多边形(图5)。一个有趣的问题是：

当 $n$ 为何值($n\geqslant 3$)时，存在格正多边形？

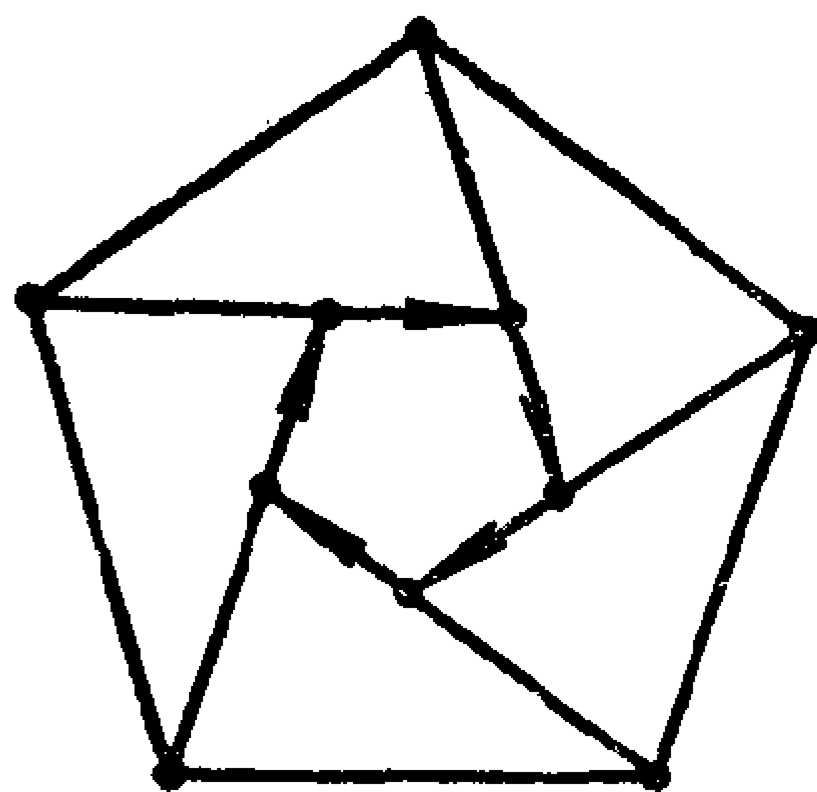
在 Scherrer 的一篇天才的文章 [25] 中给出稍微使人吃惊的答案. Hadwiger, Debrunner 和 Klee 的《平面组合几何学》一书对此作了叙述。

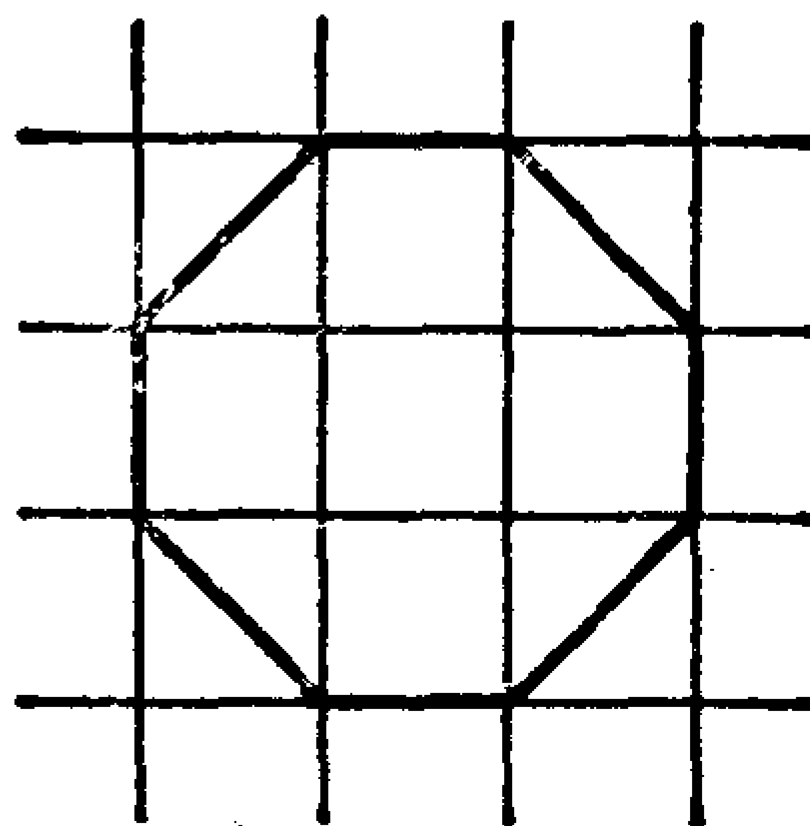
任意一个格多边形边长的公式为 $S=\sqrt{p^2+q^2}$，其中 $p,q$ 是整数。根据面积的讨论立即可知等边三角形不是格多边形。这是因为，如果格等边三角形的边长为 $S$，则面积应为 $\sqrt{3}S^2/4$；其中 $S^2$ 是一个整数。又由三角形面积的行列式公式，格三角形的面积必须是有理数，这是一个矛盾。因此格等边三角形不可能存在.同样的论证也排除了格正六边形。

接着考虑 $n\geqslant 5$，$n\neq 6$ 的情况。假定存在格正 $n$ 边形，设顶点为 $P_1,P_2,\cdots,P_n$ 的多边形是其中尺寸最小的一个。分别把这些顶点平移格向量 $\overrightarrow{P_2P_3},\overrightarrow{P_3P_4},\cdots,\overrightarrow{P_1P_2}$，得到的仍然是一个格正 $n$ 边形(见图6)，但是它包含在原先的格正 $n$ 边形的内部，这与极小性假定相矛盾。因此正方形是唯一的格正多边形。

Parsons 和 Truran [20] 给出不存在格等边三角形的循环证明. Klamkin 在[14]中曾征求、并且收到了不存在以等边三角形格为顶点的正方形的证明. Ball [1] 指出对于整数格能构造凸等边格 $n$ ($n\geqslant 3$)边形的充要条件是 $n$ 为偶数. Honsber-

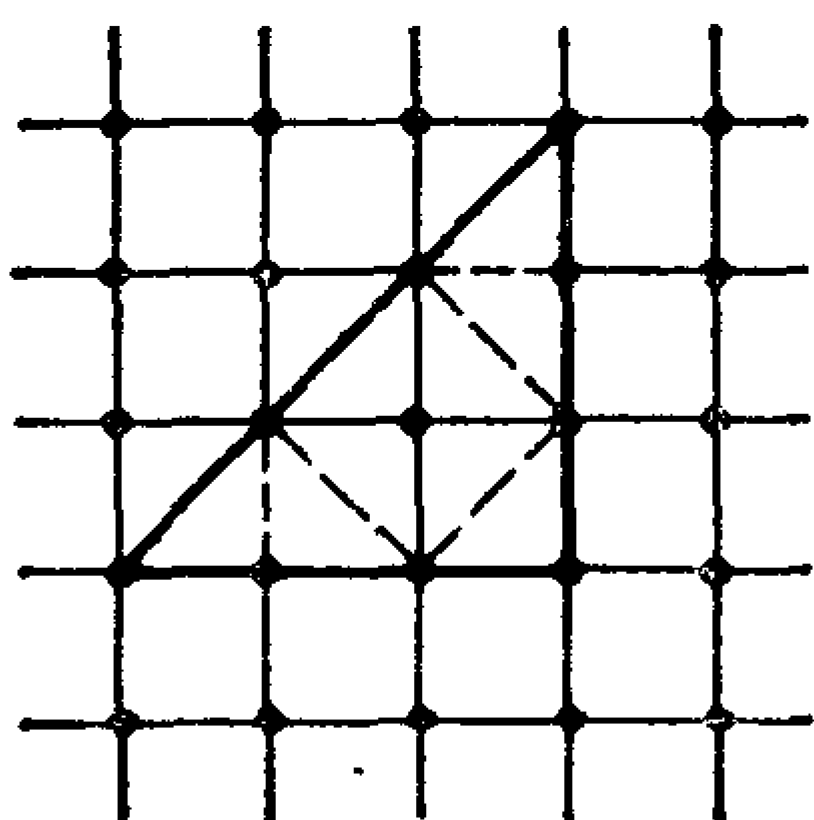
ger[13] 证明只当 $n=4$ 和 $n=8$（见图7）时，等角格 $n$ 边形才存在。关于格多边形的内角也有一些已知的结果（见文献[8]，[18]）。


图 6


图 7

### 3. 凸多边形

考虑一个凸的格多边形，在边界上有 $B$ 个格点，在内部有 $I$ 个（$I\geqslant 0$）格点。如果让内部格点数 $I$ 保持不变，而让边上的格点数 $B$ 增加来改变多边形，并且使它保持凸性，最终可以达到这样一种状态：如果再让 $B$ 增大就会破坏多边形的凸性。现在对于有给定的 $I$（$I\geqslant 0$）的凸格多边形，命 $B$ 就是这样的边界格点数的上界。


图 8

Scott[26] 和 Coleman[2] 证明了对于三角形有 $B\leqslant 2I+7$，且等号仅对图8中的三角形（或与之等价的三角形）成立。令人惊奇的是这样一个简单的结果，在如此长的年代里一直未被人发现。Coleman 对凸格 $n$ 边形还作了有趣的推测：$B\leqslant 2I+10-n$。

当 $n=3$ 时这已经是最好的估计式。但当 $n$ 增大时，这个数值可能太大了。Ehrhart[4] 证明：任意一个边数≥5的凸格多边形必包含一个内格点。

Wills[31]， Weaver[30] 和 Scott[27] 已得到关于凸格多边形的进一步结果。

## 4. 非简单多边形

Ehrhart[4] 给出 Pick 定理的一个简单推广，以适用于具有"多边形洞"的多边形区域。Reeve[22] 给出了一个更有趣的在非简单多边形情形的推广。对于多边形 $P$ 假定：

(a) 如果 $P$ 的两条边相交，它们的交点是每条边的一个顶点，因而是格点；

(b) 每一个边界点属于包含在 $P$ 中的一个非退化三角形；

(c) "由 $P$ 所界定的面积"是指从"$P$ 的外部"出发、经过奇数次跨越边界所能达到的区域的面积之和(图 9)。

现在有

$$A(P)=\frac{1}{2}B+I+k,$$

其中

$$k=-\chi(P)+\frac{1}{2}\chi(\partial P),$$

$$\chi(P)=V-E+F,$$
$$\chi(\partial P)=V-E.$$

函数 $\chi(P)$ 和 $\chi(\partial P)$ 分别是区域和边界的 Euler 示性数；$V$，$E$ 和 $F$ 分别是顶点数、边数和实际的面数。例如：

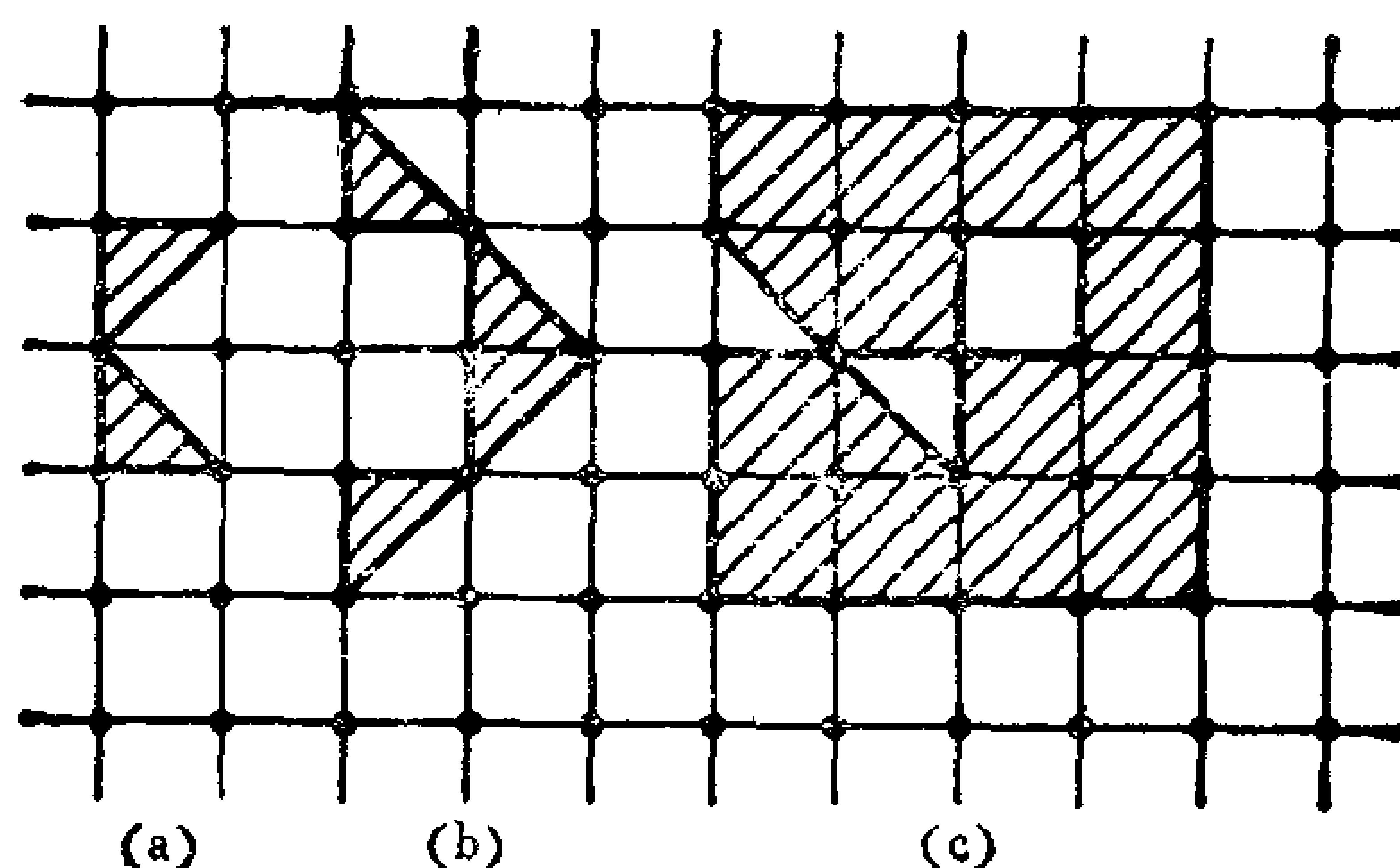



图 9

| | $B$ | $I$ | $\chi(P)$ | $\chi(\partial P)$ | $k$ | $A(P)$ |
|---|---|---|---|---|---|---|
| 图7 | 8 | 4 | 1 | 0 | $-1$ | 7 |
| 图9(a) | 5 | 0 | 1 | $-1$ | $-\frac{3}{2}$ | 1 |
| 图9(b) | 8 | 0 | 1 | $-2$ | $-2$ | 2 |
| 图9(c) | 22 | 3 | $-1$ | $-2$ | $-4$ | 14 |

由于 $V=B$，若把 $\chi(P)$，$\chi(\partial P)$的表示式代入 $A(P)$的公式中，我们便得到 Pick 定理的有趣的推广：

$$A(P)=\frac{1}{2}E+I-F.$$

对于简单多边形，$E=B$，$F=1$，上式就成为Pick 定理。Hadwiger 和 Wills [9]，Rosenholtz [23] 已经研究过这一面积公式。

## 5. 高维模拟

至今我们已在四个方面考虑过格多边形(图10)。现在自

然要问这些想法能否推广到高维情形。

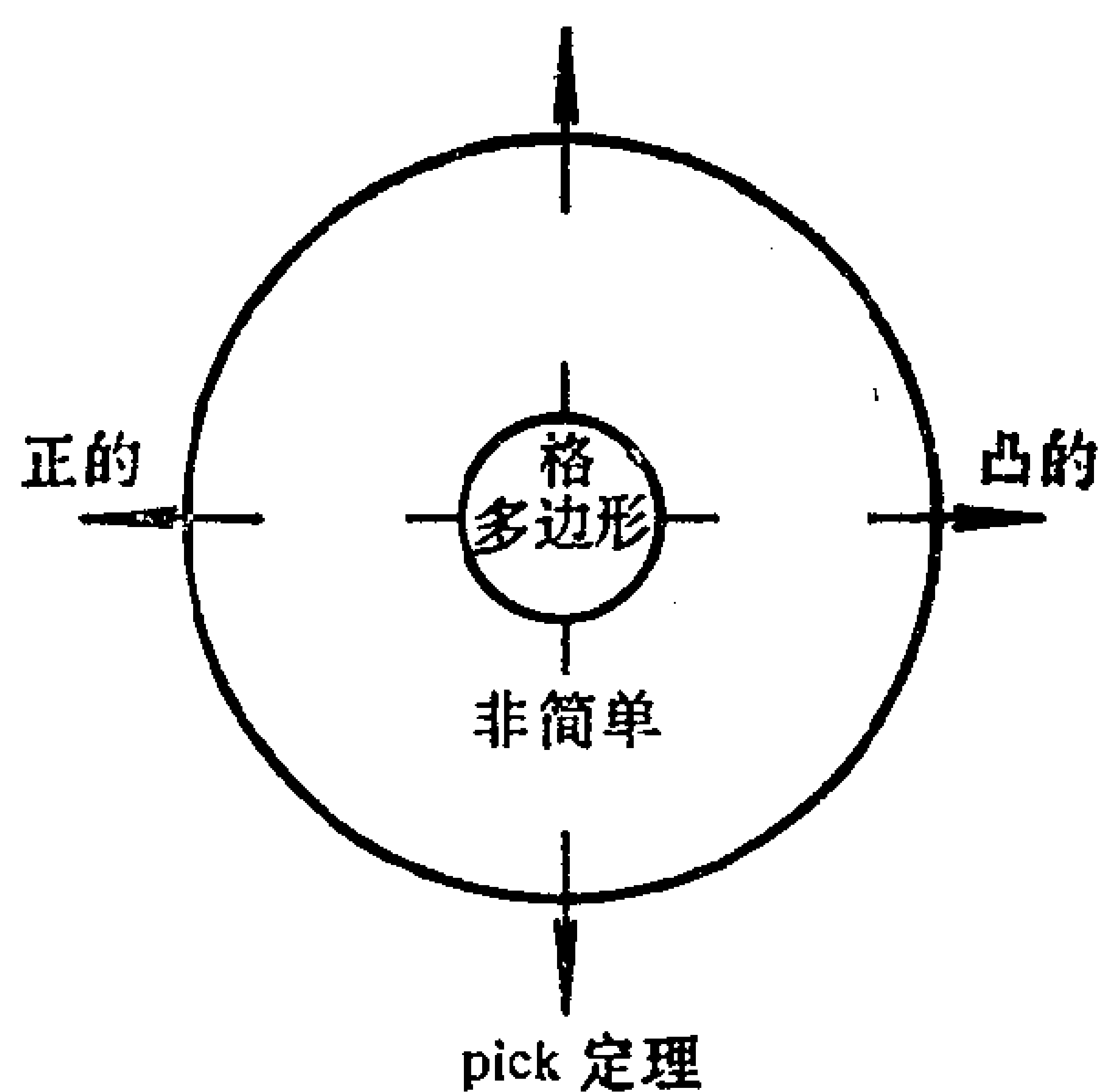



图 10

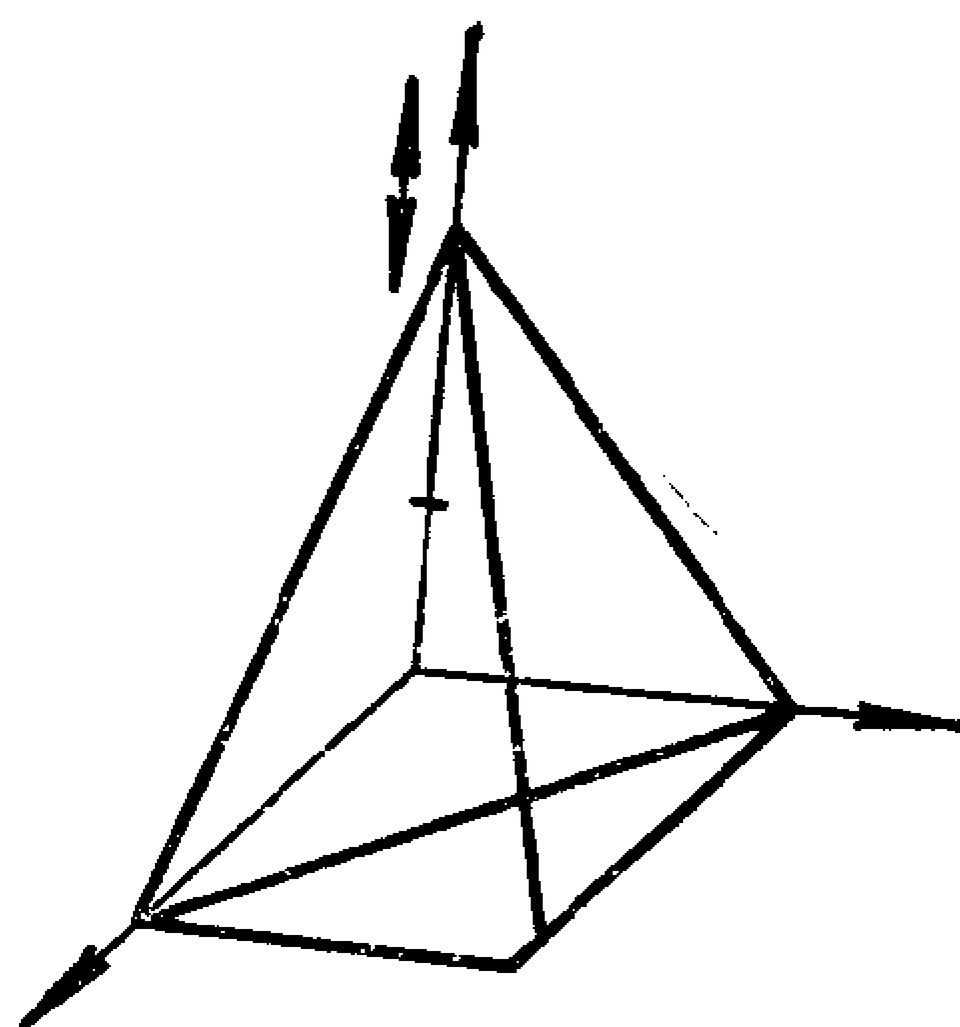


图 11

有一个反例说明在三维空间中与 Pick 定理类似的公式一定不会成立。考虑画在图 11 中的以(1,0,0),(0,1,0),(1,1,0)和(0,0,k) ($k\in \mathbf{Z}^+$) 为顶点的格四面体。对于这个四面体 $B=4$，$I=0$，体积随 $k$ 的选取而变化，因此不可能用 $B$ 和 $I$ 来表示格四面体的体积。

Reeve[22] 用一种巧妙的方式克服了这一困难。让 $L$ 表示整数格，$L_n$ ($n\in \mathbf{Z}^+$) 表示点为 $x/n$ 的子格，其中 $x\in L$。例如 $L_2$ 是点 $(a/2,b/2,c/2)$ 的集合，其中 $a,b,c$ 是整数。如果 $P$ 是一个格多面体，在其边界和内部总共包含 $L$ 的 $T$ 个格点，包含 $L_n$ 的 $T_n$ 个格点，在其边界上包含 $L$ 的 $B$ 个格点，包含 $L_n$ 的 $B_n$ 个格点，那么 $P$ 的体积 $V(P)$ 为：

$$2(n-1)n(n+1)V(P)=2(T_n-nT)-(B_n-nB).$$

正如 Reeve 所说的那样，这种公式的存在性或许比公式

本身更有趣。Reeve 给出了他的结果包括非简单多面体在内的一个推广，并且猜测在四维情形成立的一个公式。MacDonald[17] 由此建立了对于任意维数的格多面体的一般结果。

格正多边形的研究能从两个方面进行推广：(a) 可以问一个格正多边形能否被嵌入到足够高维的整数格中去？Klamkin 和 Chrestenson[15] 发现的一个充要条件是多边形具有3,4或6个顶点。(b) 考虑在三维空间中格正多面体的存在性。Ehrhart[5] 证明正二十面体和正十二面体不可能是格正多面体；其它三个正多面体能容易地被适当安置在格点上(图12)。和正方形一样，立方体既能够“正”的也能“斜”的放置在整格点上。Sárközy[24] 确定了对于 $n$ 的不同值，$n\times n\times n$立方体的所有可能的放法。在文献[28]中 Scott 证明恰有三个半正多面体可以作为格多面体。

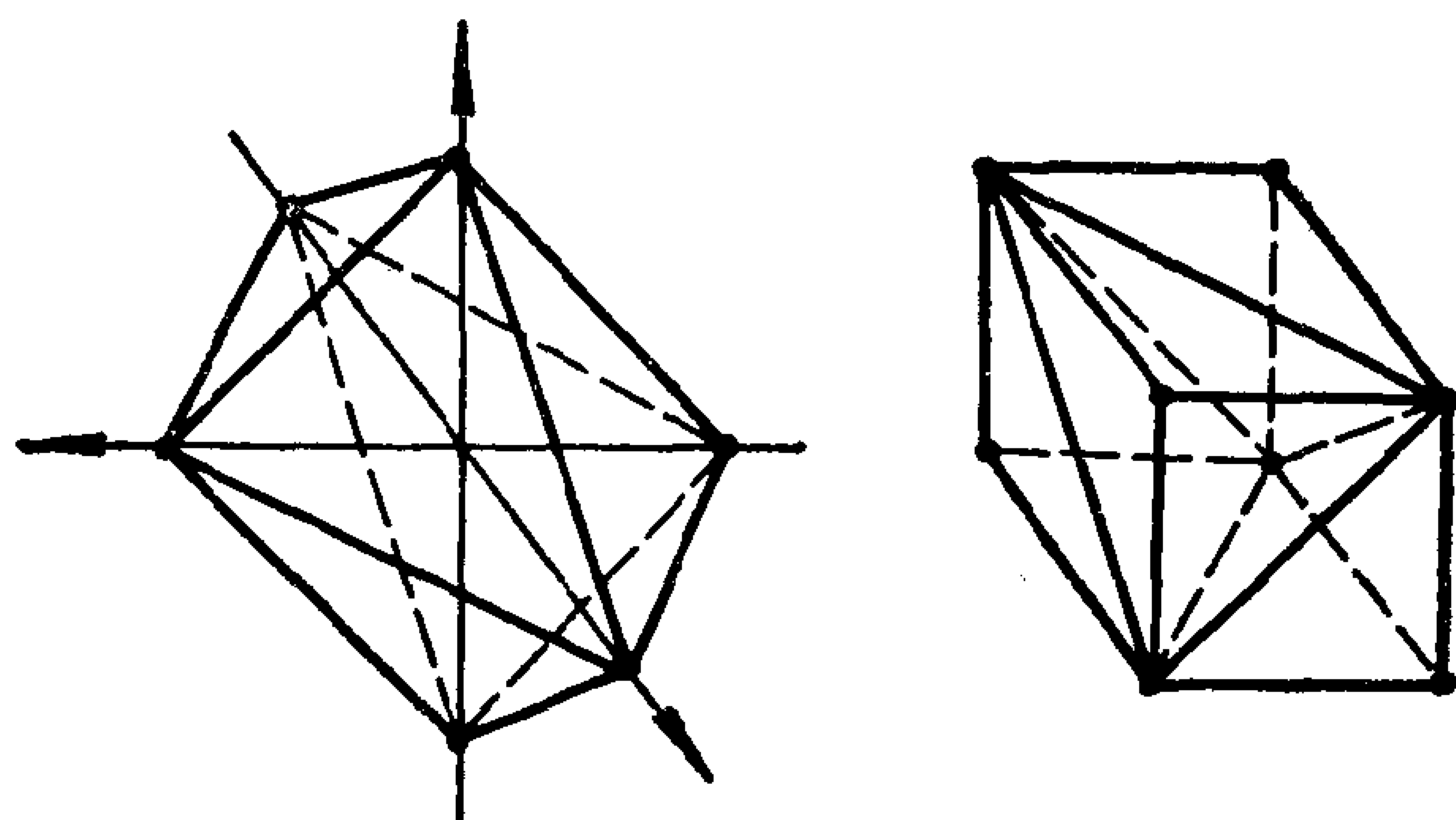


图 12

凸的格多边形的不等式 $B\leqslant 2I+7(I>0)$ 已被证实难于

推广到高维情形。Hensley[11]证明了对于一般维数中的格多面体类似不等式的存在性，Zaks, Parles, Wills[34]和Wills[32]得到一部分结果。这表明这是一个困难的问题。

通过观察 Reeve 引进子格的概念能够用来得到平面情形的 Pick 定理的另一种形式。对于平面上的子格 $L_n, L_m$，关于格多边形 $P$ 分别有：

$$A(P)=\frac{1}{n^2}\left\{\frac{1}{2}B_n+I_n-\chi(P)+\frac{1}{2}\chi(\partial P)\right\},$$

$$A(P)=\frac{1}{m^2}\left\{\frac{1}{2}B_m+I_m-\chi(P)+\frac{1}{2}\chi(\partial P)\right\},$$

因为 Euler 示性数是拓扑不变量，因此 $\chi(P)$ 和 $\chi(\partial P)$ 与格的选取无关。消去 $\chi(P), \chi(\partial P)$ 得到

$$A(P)=\frac{1}{m^2-n^2}\left\{\frac{1}{2}(B_m-B_n)+(I_m-I_n)\right\},$$

特别当 $n=1$（恰好是整数格）和 $m=2$ 时，有

$$A(P)=\frac{1}{3}\left\{\frac{1}{2}(B_2-B_1)+(I_2-I_1)\right\}.$$

这个公式对所有格多边形 $P$ 都成立。

通过这些观察我们认识到有许多有趣的数学问题正待解决，关键常常在于能够提出恰当的问题。

## 参 考 文 献

[1] D. G. Ball, The Construction of Regular and Equilateral Polygons on a Square Pinboard, Math. Gazz., 57 (1973), 119—122.

[2] D. B. Coleman, Stretch, A Geoboard Game, *Math. Mag.*, 51 (1978), 49—54.

[3] H. S. M. Coxeter, Introduction to Geometry, John Wiley (1961).

[4] E. Ehrhart, Propriètés Arithmogeometriques des Polygones, Comptes Rendus, 241 (1955), 686—689.

[5] ______, Sur les Polygones et les Polyédres Réguliers Entiers, L'Enseignement Math., 5 (1959). 81—85.

[6] W. W. Funkenbusch, From Euler's Formula to Pick's Formula Using an Edge Theorem, *Amer. Math. Monthly*, 81 (1974), 647—648.

[7] R. W. Gaskell, M. S. Klamkin, and P. Watson, Triangulations and Pick's Theorem, *Math. Mag.*, 49 (1976), 35—37.

[8] H. Hadwiger, H. Debrunner, and V. Klee, Combinatorial Geometry in the Plane, Holt, Rinehart and Winston (1964).

[9] H. Hadwiger and J. M. Wills, Neuere Studien über Gitterpolygone, *J. für* Mathematik, 280 (1973), 61—69.

[10] G. Haigh, A 'Natural' Approach to Pick's Theorem, Math. Gazz., 64 (1980), 173—177.

[11] D. Hensley, Lattice Vertex Polytopes with Interior Lattice Points, *Pac. J. Math.*, 105 (1983), 183—192.

[12] R. Honsberger, Ingenuity in Mathematics, New Mathematics Library, vol 23, MAA (1970).

[13] ______, Mathematical Gems, Two-Year College Mathematics Journal (January 1982), 36—44.

[14] M. Klamkin, Froblem 709, Elemente der Math., 30 (1975), 14—15.

[15] M. Klamkin and H. E. Chrestenson, Polygon Imbedded in a Lattice, *Amer. Math. Monthly*, 70 (1963), 447—448.

[16] A. Liu, Lattice Points and Pick's Theorem, *Math. Mag.*, 52 (1979) , 232—235.

[17] I. G. MacDonald, The Volume of a Lattice Polyhedron, *Proc. Camb. Phil. Soc.*, 59 (1963),719—726.

[18] A. Makowski, Angies of a Parallelogram with Vertices in Lattice Points, Elemente der Math.24 (1969),114—115.

[19] I. Ni,en and H. S. Zuckerman, Lattice Points and Polygonal Area, *Math. Mag.*, 40 (1967),1195—1200.

[20] R. B. Parsons and J. M. Truran,Equilateral Triangles on Geoboards, Math. Gazz., 54 (1970),53—54.

[21] G. Pick, Geometrisches zur Zahlenlehre, Sitzungsber. Lotos. Prag., 19 (1900),311—319.

[22] J. E. Reeve, On the Volume of Lattice Polyhedra, *Proc. Lond. Math., Soc.* 7 (1957),378—395.

[23] I. Rosenholtz, Calculating Surface Areas from a Blueprint, *Math. Mag.*, 52 (1979),252—256.

[24] A. Sárközy, Lattice Cubes in 3-space, (Hungarian) , Mat. Lapok., 12 (1961),232—245.

[25] W. Scherrer, Die Einlagerung eines Regularen Vielecks in ein Gitter, Elemente der Math., 1 (1946),97—98.

[26] P. R. Scott, On Convex Lattice Polygons, Bull. Austral. Math. Soc., 15 (1976),395—399.

[27] ______, An Inequality for Convex Lattice Polygons, *Math. Mag.*, 52 (1979),239—240.

[28] ______, Equiangular Lattice Polygons and Semiregular Lattice Polyhedra, The College Mathematics Journal, (to appear) .

[29] D. E. Varberg,Pick's Theorem Revisited,*Amer. Math. Monthly*, 92 (1985),584—587.

[30] C. S. Weaver, Geoboard Triangles with One Interior

Point, Math. Mag., 50 (1971), 92—94.

[31] J. M. Wills, Über konvexe Gitterpolygone, Comm. Math. Helvetici, 48 (1973), 188—194.

[32] ______ ,On an Analogue to Minkowski's Lattice Point Theorem, The Geometric Vein, Springer (1982).

[33] A. M. Yaglom and I. M. Yaglom, Challenging Mathematical Problems with Elementary Solutions, vol. 2, Holden-Day (1964).

[34] J. Zaks, M. A. Perles, and J. M. Wills, On Lattice Polytopes Having Interior Lattice Points, Elemente der Math., 37 (1982), 44—46.

(阮培文译，陈维桓校)

# 关于抛物线反射性质的证明①

ROBERT C. WILLIAMS

抛物线的一个重要性质是：从抛物镜的焦点处发射出的光线，经镜面反射后成为平行于抛物线轴的光束。这一事实等价于：设$P$是抛物线上的任意一点。用直线连接抛物线的焦点$F$和点$P$并过$P$作平行于抛物线轴的直线$l_1$，则它们与抛物线在$P$点的切线$l_2$所夹的角$\alpha$和$\beta$相等(见图1)。

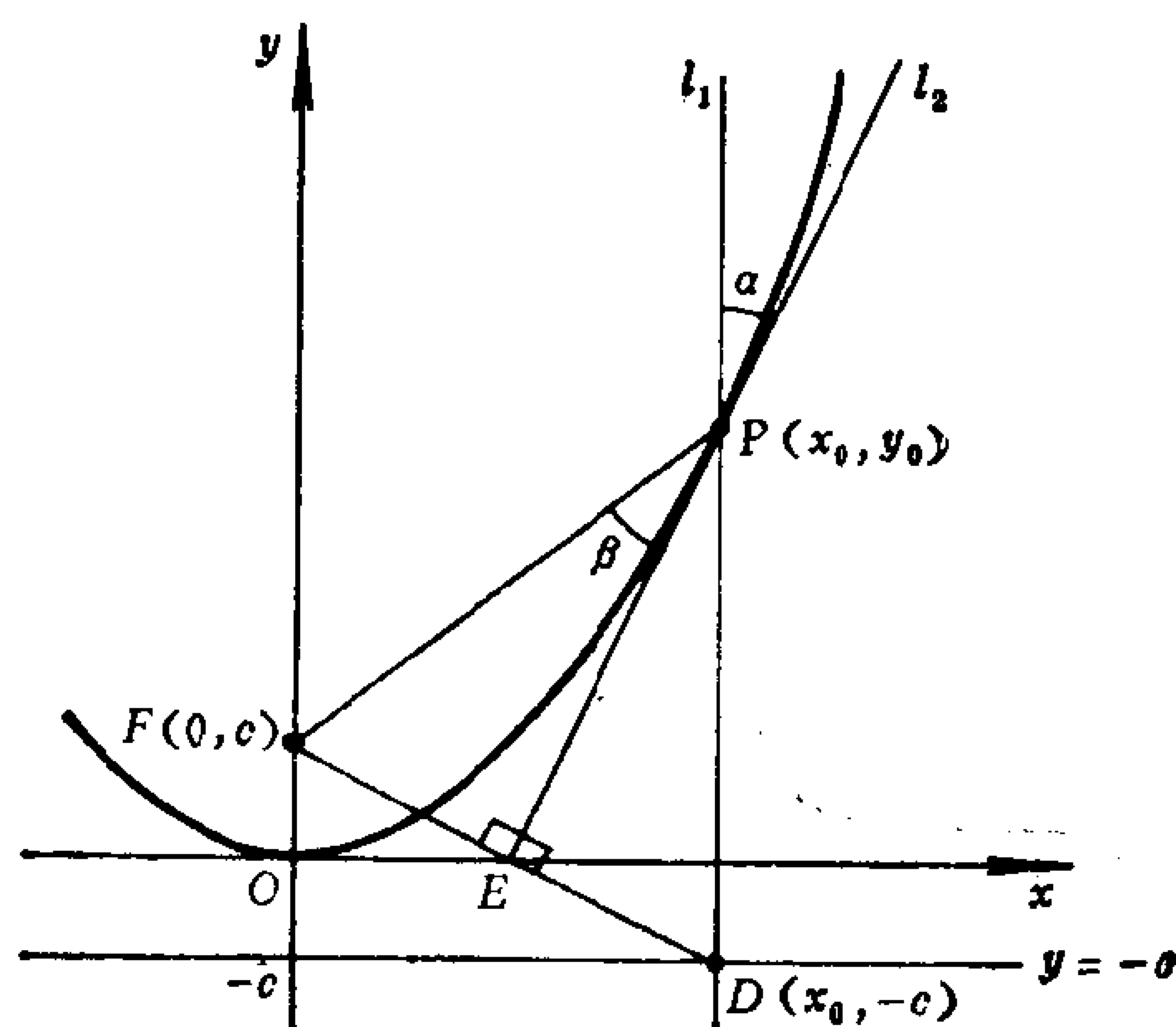



图 1

① A proof of the reflective property of the parabola, *Amer. Math. Monthly*, 94(1987), 667—668.

在我所知道的所有微积分教科书中，这条性质的证明（或者指导学生去证明）都是通过麻烦的三角形分析去证明 $\mathrm{tg}\alpha=\mathrm{tg}\beta$ 来完成的。本文给出一个更接近几何直观的简短证明，其中用了一点微积分。这一证明并不是新的（见文献[1]，P.94—95），但它确实鲜为人知。

如图所示，设抛物线的方程为 $y=x^2/4c$，焦点是 $F(0,c)$，准线为 $y=-c$。$P(x_0,y_0)$ 是抛物线上一点，$l_1$ 是过 $P$ 点且平行于抛物线轴的直线，$l_2$ 是过 $P$ 点且与曲线相切的直线。$D(x_0,-c)$ 是 $l_1$ 与准线的交点，$E$ 是 $l_2$ 与直线 $FD$ 的交点。则直线 $l_2$ 的斜率是 $\frac{x_0}{2c}$（利用微积分）而 $FD$ 的斜率是 $\frac{-2c}{x_0}$，因此它们互相垂直。因为 $|\overline{FP}|=|\overline{PD}|$，故直角三角形 $FEP$ 和 $DEP$ 全等。从而有 $\beta=\angle FPE=\angle DPE=\alpha$。

容易验证 $E$ 点在 $x$ 轴上。

## 参 考 文 献


[1] C.Smith, An Elementary Treatise on Conic Sections, Macmillan, London, 1885.


（刘　勇译，朱学贤校）

# 代数基本定理的证明①

J.L. Brenner, R.C Lyndon

代数基本定理是说每个非常数的复系数多项式在复平面内都有零点。对于这个定理，已经有许多证明了，几乎每本复分析教科书中都有它的证明。这里我们给出一个只用到初等代数知识和简单极限的证明。

**代数基本定理** 设 $P(z)$ 是一个非常数的复系数多项式，则存在复数 $z_0$ 使 $P(z_0)=0$。

下面先证几个引理。

令 $P(z)=\sum_{j=0}^{n} a_j z^j$，不妨假设 $a_0 \neq 0$，$a_n=1$。置 $A=\sum_{j=0}^{n} |a_j|$，$R=2A$。

**引理1** 如果 $|z| \geqslant R$，则

(1) $|P(z)| \geqslant |A|$，

(2) $|P(z)-z^n| \leqslant |z|^n/2$，

(3) $|\operatorname{arc}(P(z))-\operatorname{arc}(z^n)| \leqslant \pi/6$，其中 arc w 表示 $w$ 的主幅角。

**证明** (1) $|P(z)| \geqslant |z^n| - \sum_{j=0}^{n-1} |a_j||z^j|$

① Proof of the Fundamental Theorem of Algebra, *Amer. Math. Monthly*, 88(1981), 253—257.

$$\geqslant |z^{n-1}|(|z|-A+1)$$

$$\geqslant R^{n-1}(R-A)=R^{n-1}A>A.$$

(2) 因为$|z|\geqslant 2A-2$，有

$$|P(z)-z^n|\leqslant \sum_{j=0}^{n-1}|a_j||z^j|\leqslant \sum_{j=0}^{n-1}|a_j||z|^{n-1}$$

$$=(A-1)|z|^{n-1}=(A-1)|z|^{-1}|z|^n$$

$$\leqslant \frac{A-1}{2A-2}|z|^n=|z|^n/2.$$

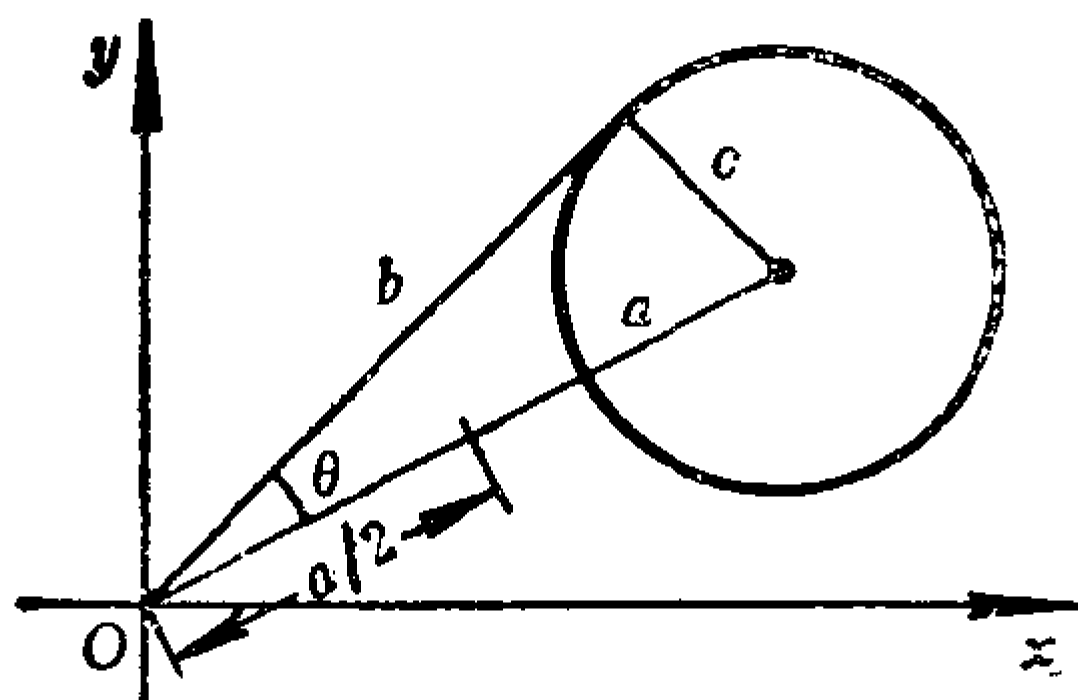



图 1

(3) 如图1所示，其中 $a=|z^n|$，$b=|P(z)|$，$c=|P(z)-z^n|$，由(2)，有 $c\leqslant \frac{1}{2}a$，而 $\theta$ 在 $b\perp c$时取极大值，因而有 $\operatorname{tg}\theta\leqslant \frac{1}{2}$，$\theta\leqslant\pi/6$。证毕。

当$w\neq 0$，$(k-1)\pi/2\leqslant \operatorname{arc} w<k\pi/2\ (k=1,2,3,4)$时，称$w$在第$k$象限，记为$Q(w)=k$。若$\frac{1}{2}|Q(z_1)-Q(z_2)|=1$，称$z_1,z_2$在相对的象限中。

**引理2** 如果$z_1,z_2$在相对的象限中，则

$$|z_1|,|z_2|\leqslant|z_2-z_1|.$$

**证明** 见图2。

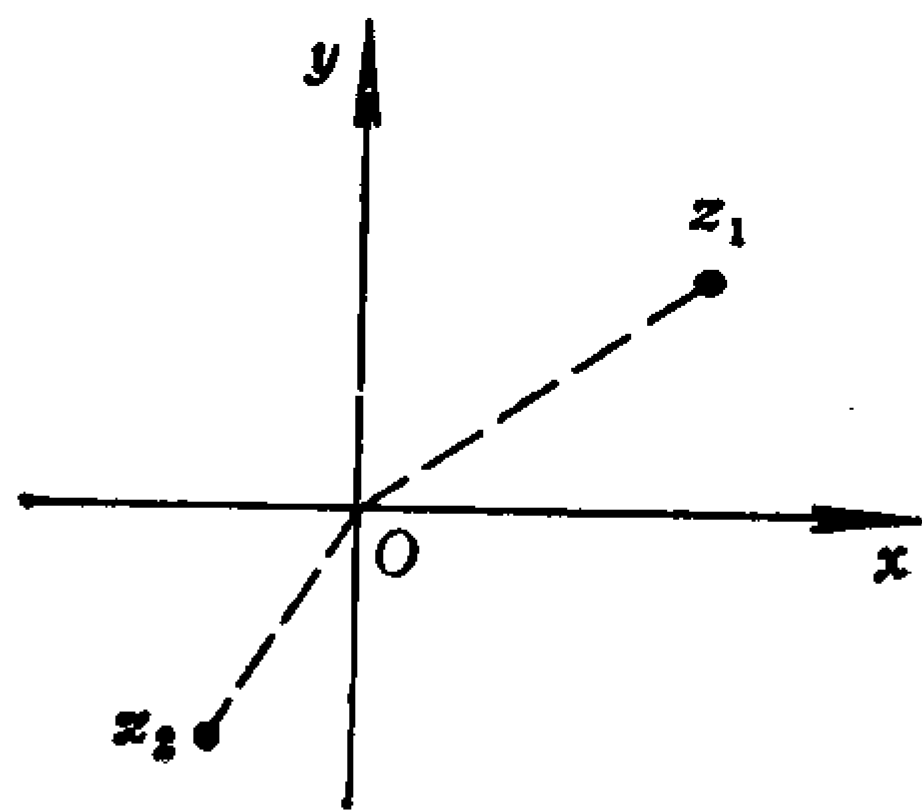



图 2

**引理3** 如果$S>0$，则存在一个只依赖于$S$和$P$的正整数$K$，使当$|z_1|,|z_2|\leqslant S$时，就有$|P(z_2)-P(z_1)|$

$\leqslant K|z_2-z_1|$。

**证明** 不妨设 $S>1$。则因为

$$P(z_2)-P(z_1)=\sum_{j=0}^{n}a_j(z_2^j-z_1^j)$$

$$=(z_2-z_1)\sum_{j=1}^{n}a_j\sum_{h+k=j-1}z_1^h z_2^k,$$

因而

$$|P(z_2)-P(z_1)|\leqslant|z_2-z_1|\sum_{j=1}^{n}|a_j|S^{j-1}$$

$$\leqslant|z_2-z_1|S^nA。$$

**定理的证明** 下面取 $K=A\{\max(1,S)\}^n$。任取一数 $\varepsilon$，使 $0<\varepsilon<1$。选取一包含 $|z|=R$ 的等边三角形 $T$ 和一正整 $S$，使 $T$ 含于圆 $|z|=S$ 中。令 $\delta=\varepsilon/K$。取 $\Delta$ 为 $T$ 中的 2-维闭集合。则由引理 3，有

(4) $z_1,z_2\in\Delta$，$|z_1-z_2|<\delta\Rightarrow|P(z_1)-P(z_2)|<\varepsilon$。

现在我们用平行于 $T$ 的边的等距线把 $\Delta$ 等分成一些全等的等边三角形 $\Delta_\nu$，而得到一网格。如图 3 所示。下面称同一个小三角形的任意两顶点为相邻点。把这个网格取得充分小使每个小三角形的边长不大于 $\delta$，这样若 $z,z'$ 是 $T$ 的边上的相邻格点，

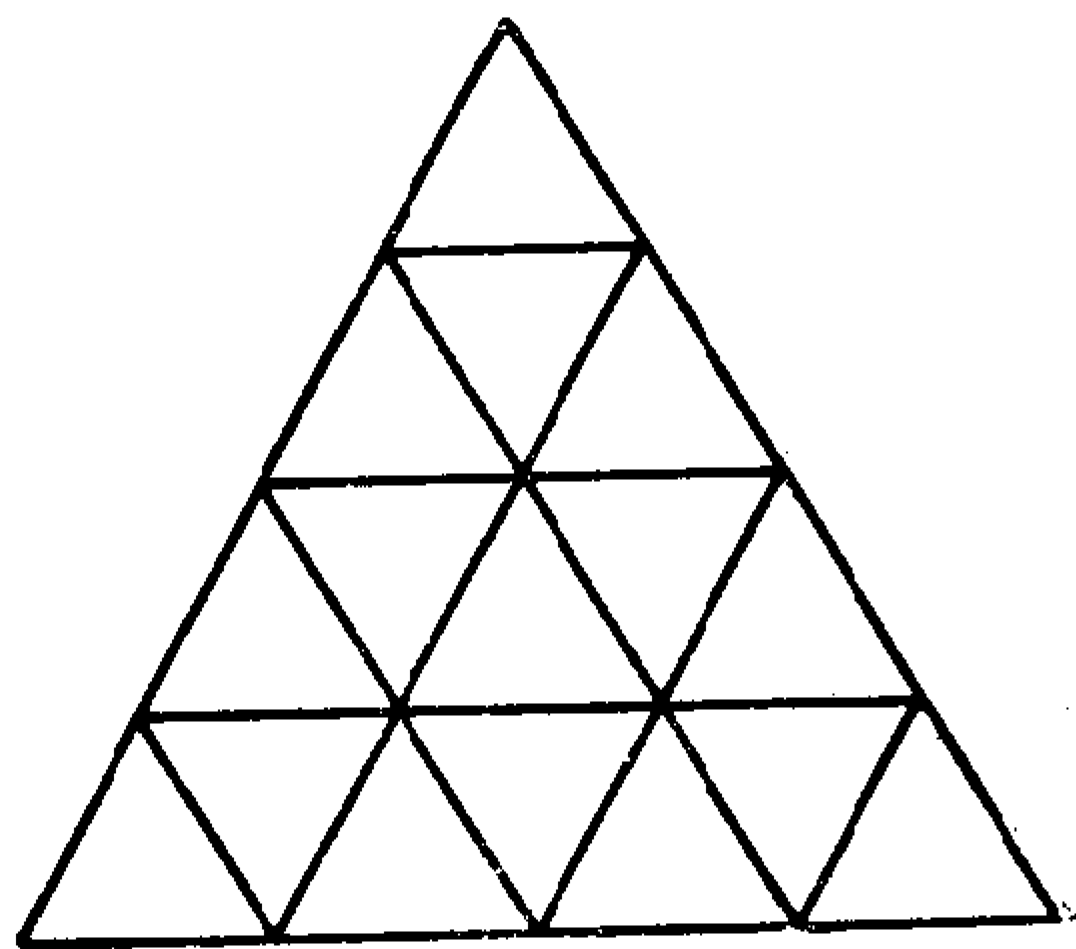


图 3

则$|\text{arc}\ z' - \text{arc}\ z| < \pi/6n$。如果$z, z'$是任意两个相邻格点，则$|z' - z| < \delta$，因此由(4)有，$|P(z') - P(z)| < \varepsilon$。

假设任意相邻点$z, z'$ 都使$w = P(z)$与$w' = P(z')$不在相对的象限中。也就是说，$[Q(w') - Q(w)] \pmod 4$的值是$-1$，0或1．下面记$[Q(w') - Q(w)] \pmod 4$为$d(w, w')$。我们将导出矛盾。

令$z_1, \cdots, z_t$是三角形$T$的边上的点按反时针方向取循环次序。令$w_j = P(z_j)$．则由于我们选择的网格充分的小，序列

$$Q(z_1^n), \cdots, Q(z_t^n) \qquad (*)$$

除重复的以外，恰好跑过圈(1,2,3,4) $n$次，由(3)，序列

$$Q(w_1), \cdots, Q(w_t) \qquad (**)$$

与序列(*)只在形为$h, \cdots, h, h', \cdots, h'$，其中$h' \equiv h + 1 \pmod 4$的(连续的)子序列处不相同，把这样的子序列记为$h, h_1, \cdots, h_m, h'$ 其中$h_i$是$h$或$h'$ 定义$D(\Delta) = \sum_{j=1}^{t} d(w_{j+1}, w_j)$，其中$w_{t+1} = w_1$。那么，我们有

$$D(\Delta) \equiv \sum_{j=1}^{t} d(w_{j+1}, w_j) = \sum_{j=1}^{t} d(z_{j+1}^n, z_j^n) = 4n.$$

令$z', z'', z$ 是某个$\Delta_\nu$的顶点，依反时针方向标号．则由假设，$Q(P(z')), Q(P(z''))$，$Q(P(z))$至多有两个相邻的值$k, k'$，即$k' \equiv k + 1 \pmod 4$，从而$D(\Delta_\nu) = d(w'', w') + d(w''', w'') + d(w', w) = 0$。因此，$\sum D(\Delta_\nu) = 0$，其中求和是对此网格中的所有三角形$\Delta_\nu$．

假设$z', z''$是网格的内部相邻点(即不在$T$的边上的点)。则它们是两个三角形$\Delta_\nu$和$\Delta_\mu$的顶点，因而$d(w'', w')$，

$d(w',w'')$将各在$D(\Delta_\nu)$，$D(\Delta_\mu)$之一中出现。由于$d(w',w'')=-d(w'',w')$，因而这两项将在求和$\sum D(\Delta_\nu)$中对消掉。因此，$\sum D(\Delta_\nu)$只剩下了由$T$的边上的点得到的项$d(w',w'')$之和，即$\sum D(\Delta_\nu)=D(\Delta)$。

由于$D(\Delta)=4n$，因此$\sum D(\Delta_\nu)=0$。矛盾。所以，存在两个相邻点$z,z'$，使$P(z),P(z')$位于相对的象限中，则由引理2，我们有

$$|P(z)|<|P(z')-P(z)|<\varepsilon.$$

这样我们就证明了，对任意$\varepsilon>0$，如果$z,z'$是两个相邻格点，都有

$$|z-z'|<\varepsilon/K$$

和

$$|P(z)|<|P(z')-P(z)|<\varepsilon.$$

由复平面的连续性和柯西收敛准则，我们有，当$\varepsilon\to0$时$z\to z_0$且$f(z)\to0$，即$f(z_0)=0$。这就证明了我们的定理。

（李才恒编译，徐明曜校）

# 叠二项式系数①

SOLOMON W. GOLOMB

## 1. 引言

研究运算的叠置是数学上的传统课题，当这些运算是非交换和不可结合的时候，它们通常揭示出许多十分有趣的性质。通常考虑形如$c^{b^a}$这种指数的运算，但对于叠二项式系数

$$\binom{\binom{c}{b}}{a}$$

的讨论却不多见。这个表达式有一个自然的组合学的解释："先从 $c$ 个元素中选取 $b$ 个元素构成一个$b$-元子集，再从所有这些 $b$- 元子集中选取 $a$ 个的方法数"。即从 $c$ 个元素的集合选取 $a$ 个 $b$- 元子集的方法数。特别地，当我们试图进一步考虑叠置

$$\left[\binom{\binom{d}{c}}{b}\right]$$

的时候，也许是因为记号太笨拙的缘故，这个题目一直不被

① Iterated binomial coefficients, *Amer. Math. Monthly*, 87 (1980), 719—727.

人注意（旧的组合记号如

$$C_a^{C_b^c},\quad C_a^{C_b^{C_c^d}},\cdots$$

比竖式记号好不了多少!)。

为方便起见，我们给出一种更便于书写的记号。归纳定义$(a_1;a_2;\cdots;a_k)$如下：

$$(a_1)=a_1,\qquad (a_1;a_2)=\binom{a_1}{a_2},$$

$$(a_1;a_2;\cdots;a_{k-1};a_k)=((a_1;a_2;\cdots;a_{k-1});a_k).$$

因此

$$(a_1;a_2;a_3)=\left(\begin{matrix}\binom{a_1}{a_2}\\ a_3\end{matrix}\right).$$

当$k\leqslant3$时，由于我们比较熟悉"竖式记号"且它的意义也直观，因而在使用"水平记号"的同时也使用"竖式记号"。

本文研究恒等式，不等式，方程，可除性，极值问题和含有叠二项式系数的简化公式。我们不打算作过细的论述，而只是将它们作为进入丰富的新领域的一种初步的研究。

## 2. 第一恒等式

**定理1**

$$\left(\begin{matrix}\binom{n}{2}\\ 2\end{matrix}\right)=3\binom{n+1}{4}.$$

**证明** I（代数法）。

$$(n;2;2)=\binom{\frac{n^2-n}{2}}{2}=\frac{1}{2}\left(\frac{n^2-n}{2}\right)\left(\frac{n^2-n-2}{2}\right)$$

$$=\frac{1}{8}(n+1)(n)(n-1)(n-2)=3\binom{n+1}{4}.$$

Ⅱ(组合法). $(n;2;2)$是从 $n$ 个物体的集合里选取2-元子集对的方法数。如果从 $n$ 个物体的集合里选取4个元素(比方说 $a,b,c,d$),那么有3种方法把它们排列为 2-元子集对($ab$ 对 $cd$,$ac$ 对 $bd$,$ad$ 对 $bc$)。因为在每对 2-元子集中可能有一个公共元素,所以我们得不到$(n;2;2)=3(n;4)$。设想在 $n$ 张纸牌中添加一张"百搭",现取出其中的 4 张,如果是 4 张普通的牌,那么有 3 种方法把它们排列为2-元子集对。如果是 3 张普通牌(比方说 $a,b,c$)加一张"百搭",那么(也!)有 3 种方法形成 2-元子集对($ab$ 对 $ac$,$ab$ 对 $bc$,$ac$ 对 $bc$)。因此有$(n;2;2)=3(n+1;4)$。

定理 1 是二项式系数的一个简化公式(从 3-层到 2-层的二项式系数),也包含了一个不等式($(n;2;2)>(n;4)$,$n\geqslant 3$),同时也给出了一个可除性结果($(n+1;4)$整除$(n;2;2)$,其中的两个表达式都被看作有理数域上以 $n$ 为变量的多项式)。本文将进一步研究这些内容。

定理 1 的内容首先出现在[1]中。

## 3. 可除性结果

**定理2** 对所有的 $k\geqslant 1$,有

$$\binom{n+k-1}{2k} \text{ 整除 } \begin{bmatrix} \binom{n}{2} \\ \binom{k}{2}+1 \end{bmatrix},$$

其中的两个表达式都被看成是有理数域上以 $n$ 为变量的多项式。

**证明** 注意到

$$\begin{pmatrix} \binom{n}{2} \\ a \end{pmatrix} = \frac{(n^2-n)(n^2-n-2)(n^2-n-4)\cdots(n^2-n-2a+2)}{2^a \cdot a!}.$$

当 $a=1,2,4,7,\cdots,\binom{k}{2}+1,\cdots$时，对应的二次因式为

$$n^2-n=n(n-1), \qquad n^2-n-2=(n+1)(n-1),$$
$$n^2-n-6=(n+2)(n-3),$$
$$n^2-n-12=(n+3)(n-4),\cdots,$$
$$n^2-n-2\binom{k}{2}=(n+k-1)(n-k),\cdots.$$

这表明

$$\begin{bmatrix} \binom{n}{2} \\ \binom{k}{2}+1 \end{bmatrix}$$

的多项式因子中包含有

$$\prod_{j=-k}^{k-1}(n+j)=(2k)!\binom{n+k-1}{2k}.$$

由于常数因子$(2k)!$不影响有理数域上多项式的可除性，结论得证。

**注** 定理2中的两个多项式的表达式仅当$k=1$和$k=2$时有相同的次数。当$k>2$时，$(n;2;(k;2)+1)$含有不可分解的二次因式，但它们并不包含在$(n+k-1;2k)$之中。

**4. $(c;b;a)$与$(c;(b;a))$的比较**

设$a,b,c$为正整数，我们要寻求"相关"方程

$$\binom{\binom{c}{b}}{a}=\binom{c}{\binom{b}{a}}$$

的所有解。

**定理3** 除去$1<a<b<c$的情形之外，

$$\binom{\binom{c}{b}}{a}=\binom{c}{\binom{b}{a}}$$

的所有解是下面的5种类型之一：

(i) 如果$a=1$，则方程对所有的$b,c$都有解。

(ii) 如果$a>b$，则方程有解当且仅当$a=\binom{c}{b}$。

(iii) 如果$b>c$，则方程有解当且仅当$a<b$。

下面给出的两种类型的解涉及到$1<a\leqslant b\leqslant c$的情形，其中

(iv) 如果$a=b$，方程有解当且仅当$c=a+1$。

(v) 如果$b=c$，方程有解当且仅当$a<c-1$。

**证明** (i) 因为$a=1$，所以$\binom{\binom{c}{b}}{1}=\binom{c}{b}=\binom{c}{\binom{b}{1}}$。

(ii) 如果 $a>b$，则 $\binom{b}{a}=0$，$\binom{c}{\binom{b}{a}}=\binom{c}{0}=1$，于是，

$$\binom{\binom{c}{b}}{a}=1 \text{ 当且仅当 } a=\binom{c}{b}.$$

(iii) 如果 $b>c$，就有

$$\binom{\binom{c}{b}}{a}=\binom{0}{a}=0,$$

则

$$\binom{c}{\binom{b}{a}}=0 \text{ 当且仅当 } \binom{b}{a}>c.$$

因为 $c\geqslant 1$，$b>c$，这就排除了 $a\geqslant b$，所以有 $a<b$。

(iv) 如果 $a=b$，此时有

$$\binom{c}{\binom{b}{a}}=\binom{c}{1}=c,$$

那么

$$\binom{\binom{c}{b}}{a}=\binom{\binom{c}{a}}{a}=c$$

当且仅当 $a=1$ 或 $a=c-1$。因为 $a=1$ 已包含在(i)中，故唯一的"新"解是 $a=b=c-1$。

(v) 如果 $b=c$，$a>1$，此时有 $\binom{\binom{c}{b}}{a}=\binom{1}{a}=0$，那么

$$\binom{c}{\binom{b}{a}} = \binom{c}{\binom{c}{a}} = 0$$

当且仅当$\binom{c}{a}>c$；即当且仅当 $1<a<c-1$。

如上所述，如果 $1<a<b<c$ 不成立，定理中的相关方程没有其它类型的解。在这种情况下，对 $a,b$和 $c$ 的“所有”选法凭经验似乎可得

$$\binom{\binom{c}{b}}{a} > \binom{c}{\binom{b}{a}}.$$

实际上，表面现象往往靠不住，尽管当$(c;b;a)<10^{18}$时对上述不等式没有反例是事实，但有一种情况，其中的反例超过了我们所提供的例子。这种情况就是：

**定理4** 对满足$\binom{b}{a}>ab$的固定的 $a,b$，存在常数 $c_0=c_0(a,b)$，使得对一切 $c\geqslant c_0$ 有

$$\binom{\binom{c}{b}}{a} < \binom{c}{\binom{b}{a}}.$$

**证明** 把 $n$ 当作变量，同时固定 $k$，则

$$\binom{n}{k}\sim\frac{n^k}{k!}\ (\text{当 } n\to\infty\text{时}).$$

于是

$$\binom{\binom{c}{b}}{a}\sim\binom{\frac{c^b}{b!}}{a}\sim\frac{c^{ab}}{a!\,(b!)^a}\ (\text{当 } c\to\infty\text{时}),$$

但

$$\binom{c}{\binom{b}{a}} \sim \frac{c^{\binom{b}{a}}}{\binom{b}{a}!}\text{（当 } c\to\infty \text{时）。}$$

根据假设$\binom{b}{a}>ab$，所以$c^{\binom{b}{a}}$比$c^{ab}$增长得快；而且即使$\binom{b}{a}!>a!(b!)^a$，它们也只是些对充分大的$c$不起什么影响的常数。因此当$c$充分大时，$\dfrac{c^{\binom{b}{a}}}{\binom{b}{a}!}>\dfrac{c^{ab}}{a!(b!)^a}$。

注 1　$\binom{b}{a}>ab$ 的“最小的”情况包括 $15=\binom{6}{2}>2\cdot 6=12$ 和 $24=\binom{7}{2}>2\cdot 7=14$。根据大量的数值上的研究确定出最小的 $c_0=c_0(a,b)$ 是 $c_0(2,7)=75$，它还使$(c_0;b;a)$和$(c_0;(b;a))$取最小值，它们的值为

$$\left[\begin{matrix}\binom{75}{7}\\ 2\end{matrix}\right]\approx 1.97\times 10^{18}<\left[\begin{matrix}75\\ \binom{7}{2}\end{matrix}\right]\approx 2.10\times 10^{18}.$$

**注2**　对 $a=2$，$b=6$，我们得到 $c_0(2,6)=132$，有

$$\left[\begin{matrix}\binom{132}{6}\\ 2\end{matrix}\right]\approx 2.143\times 10^{19}<\left[\begin{matrix}132\\ \binom{6}{2}\end{matrix}\right]\approx 2.154\times 10^{19}.$$

**注3**　$c_0$ 的值实际是稍大于方程

$$\frac{c^{\binom{b}{a}}}{\binom{b}{a}!}=\frac{c^{ab}}{a!(b!)^{a}}$$

的解 $c=c_1$ 的值。当 $n\to\infty$ 时渐近关系式 $\binom{n}{k}\sim n^k/k!$ 中的 $\binom{n}{k}$ 可用等式 $(n,k)=\bar{n}/k!$ 来替换，其中 $\bar{n}=\text{g.m.}(n,n-1,\cdots,n-k+1)\approx n-\frac{1}{2}(k-1)$，“g.m.”是指几何平均值，它小于算术平均值 $n-\frac{1}{2}(k-1)$。令

$$\binom{c}{\binom{b}{a}}=\frac{\bar{c}^{\binom{b}{a}}}{\binom{b}{a}!},\qquad \binom{\binom{c}{b}}{a}=\frac{\hat{c}^{ab}}{a!(b!)^{a}},$$

容易证明 $\bar{c}>\hat{c}$，因此 $c_0>c_1$。

注4　当 $1<a<b<c$ 时，相关方程

$$\binom{\binom{c}{b}}{a}=\binom{c}{\binom{b}{a}}$$

是否有解仍然是一个未解决的问题。这种解的存在性似乎是十分靠不住的，但要证明这一点看来非常困难。

## 5. 一个简单不等式

对于任意由 $n$ 个物体组成的集合，其中 $n$ 充分大足以构成3-元子集，那么从中选取 3 个一组的2-元子集的方法数多

于选取 2 个一组的3-元子集的方法数。

**定理5**

$$\binom{\binom{n}{2}}{3} > \binom{\binom{n}{3}}{2}, \qquad n \geqslant 3.$$

**证明**

$$\binom{\binom{n}{2}}{3} = \frac{\dot{n}^6}{48} > \binom{\binom{n}{3}}{2} = \frac{\tilde{n}^6}{72},$$

由于 72>48 且 $\dot{n} > \tilde{n}$，所以上述不等式成立。

**注1** 上述证明实际上也给出了 当 $n \geqslant 3$ 时有

$$\binom{\binom{n}{2}}{3} > \frac{3}{2}\binom{\binom{n}{3}}{2}.$$

下一节将要用到定理 5 中的不等式。对此不等式目前还没有给出明显的纯组合学的解释.

**注2** 类似的问题如$a^{(b^c)} = (a^b)^c$（见文献[ 2 ]）.若整数 $a,b,c$ 满足$1<a<b<c$，则有严格不等式 $a^{\binom{c}{b}} > (a^b)^c$。

**注3** 定理 5 的推广是：若 $1<a<b<c$，那么

$$\binom{\binom{c}{a}}{b} > \binom{\binom{c}{b}}{a}.$$

这个结论的证明完全类似于定理 5 。特别的，如果$1<a<b$，则有 $a!(b!)^a > b!(a!)^b$。

## 6. 遍历剖分的最大值

设 $n$ 为自然数，$n$ 的一个剖分是指 $n = a_1 + a_2 + \cdots + a_k$，

其中 $a_1, a_2, \cdots, a_k$ 为正整数。我们感兴趣于使叠二项式系数 $(a_1; a_2; \cdots; a_k)$ 达到最大的那种剖分。为启发这一方法并期望得到解的形式，我们先考虑已在文献中出现过的两个相似而又较简单的问题(对定理 7 见文献[ 3 ])。

**定理6** 设 $g(n) = \max\{a_1 \cdot a_2 \cdots a_k\}$，其中对构成 $n$ 的所有剖分 $n = a_1 + a_2 + \cdots + a_k$ 中的正整数 $a_1, a_2, \cdots, a_k$ 的乘积取最大值，则 $n>1$ 时有

$$g(n) = \begin{cases} 3 \cdot 3 \cdots 3 = 3^{n/3} & \text{如果 } n \equiv 0 \pmod 3, \\ 3 \cdot 3 \cdots 3 \cdot 4 = 4 \cdot 3^{(n-4)/3}, & \text{如果 } n \equiv 1 \pmod 3, \\ 3 \cdot 3 \cdots 3 \cdot 2 = 2 \cdot 3^{(n-2)/3}, & \text{如果 } n \equiv 2 \pmod 3. \end{cases}$$

**证明** 当 $n \leqslant 4$ 时，由穷举法可得 $g(n) = n$。当 $n \geqslant 5$ 时，显然“最优化原理”成立，即

$$g(n) = \max_{1<a<n} \{a \cdot g(n-a)\}.$$

此外，使上式成立的 $a<5$，这是因为不等式 $5g(n-5) < 2 \cdot 3g(n-5)$ 表明 $n$ 有一个比包含5做为其中一部分的剖分更好的剖分；更不用说 $a>5$ 的情形了。当 $n$ 较大时，使 $ag(n-a)$ 最大的 $a$ 的值可通过比较下面三个式子：$2g(n-2)$，$3g(n-3)$，$4g(n-4)$ 而求得。为了便于比较，把每个式子叠置足够多次得到

$$2^6 g(n-12) < 3^4 g(n-12) > 4^3 g(n-12).$$

于是当 $n$ 大时，所求的 $a$ 的最好值是 $a=3$。事实上，$n>4$ 时 $g(n) = 3g(n-3)$ 定义了通解。这个通解在满足 $g(2)=2$，$g(3)=3$ 和 $g(4)=4$ 时给出了三条下降线（因为 $a$ 的“标准值”为3）.

**定理7** 设 $h(n) = \max\{a_1^{a_2^{\cdot^{\cdot^{\cdot^{a_k}}}}}\}$，其中 max 是对 $n$ 的所有

剖分 $n=a_1+a_2+\cdots+a_k$ 取的。那么当 $n>4$ 时有

$$h(n)=\begin{cases}2^{2^{\cdot^{\cdot^{2^{3^2}}}}}, & n\text{为奇数,}\\ 2^{2^{\cdot^{\cdot^{2^{3^3}}}}}, & n\text{为偶数。}\end{cases}$$

**证明** 当 $n\leqslant4$ 时，根据穷举法有 $h(n)=n$。当 $n\geqslant5$时，显然"最优化原理"成立，即

$$h(n)=\max_{1<a<n} a^{h(n-a)}.$$

此外，因为只要 $h(n-4)>2$ 成立就有 $4^{h(n-4)}<2^{2^{h(n-4)}}$.故使上式成立的 $a<4$。因此只要比较 $2^{h(n-2)}$ 与 $3^{h(n-3)}$，或比较 $h(n-2)\log 2$ 与 $h(n-3)\log 3$的大小就可以了。要使

$$h(n-2)\log 2>h(n-3)\log 3,$$

只要 $$h(n-2)/h(n-3)>\log 3/\log 2\approx1.585,$$

容易看出上式对所有的 $n>6$ 成立。于是除 $h(n)=2^{h(n-2)}$ 之外，临界情况是 $h(5)=3^2=9$ 和 $h(6)=3^3=27$。

**定理8** 设 $f(n)=\max(a_1;a_2;\cdots;a_k)$，其中最大值是在对 $n$ 的所有正整数剖分取的。则当 $n>10$ 时有

$$f(n)=\begin{cases}(5;2;2;3;3;\cdots;3), & \text{如果 } n\equiv0\pmod 3,\\ (5;2;2;2;2;3;3;\cdots;3), & \text{如果 } n\equiv1\pmod 3,\\ (5;2;2;2;3;3;\cdots;3), & \text{如果 } n\equiv2\pmod 3.\end{cases}$$

当 $n\leqslant10$ 时有

$f(1)=1,\quad f(2)=2,\quad f(3)=3,\quad f(4)=4,\quad f(5)=5,$

$$f(6)=\binom{4}{2}=6,\quad f(7)=\binom{5}{2}=10,\quad f(8)=\binom{6}{2}=15,$$

$$f(9)=\begin{bmatrix}\binom{5}{2}\\ 2\end{bmatrix}=45,\quad f(10)=\begin{bmatrix}\binom{5}{2}\\ 3\end{bmatrix}=120.$$

**证明** 当 $n\leqslant 10$ 时，$f(n)$ 的值仍可由穷举法验证。当 $n>10$ 时，“最优化原理”成立，即

$$f(n)=\max_{1<a<n}\binom{f(n-a)}{a}.$$

事实上，使上式成立的 $a<4$，这是因为由定理 1 得

$$\binom{f(n-4)}{4}<\begin{bmatrix}\binom{f(n-4)}{2}\\ 2\end{bmatrix},$$

更不用说 $a>4$ 的情形了，因此只有 $a=2$，$a=3$ 可供选择。又由定理 5 有

$$\binom{\binom{n}{2}}{3}>\binom{\binom{n}{3}}{2},$$

因此在 $f(n)$ 的表达式中“2”决不会出现在“3”的下面（这是指相对于竖式的记号）。从而，当 $n>5$ 时，$f(n)$ 的值或是 $\binom{f(n-2)}{2}$，或是 $\binom{f(n-2)}{3}$。注意到

$$\binom{\binom{m}{3}}{3}\sim\frac{m^9}{1296},\quad \begin{bmatrix}\binom{\binom{m}{2}}{2}\\ 2\end{bmatrix}\sim\frac{m^8}{128}\quad（当\ n\to\infty 时），$$

这就意味着存在 $m_0$，当 $m\geqslant m_0$ 时有

$$\binom{\binom{m}{3}}{3}>\begin{bmatrix}\binom{\binom{m}{2}}{2}\\ 2\end{bmatrix},$$

因此，对所有的 $n \geqslant n_0$，有 $f(n)=\binom{f(n-3)}{3}$.经过仔细的验证得出 $n_0=14$。由于 $n \geqslant 14$ 依赖于 $n$(模 3 )的值，故$f(n)$为 $(f(11);3;3;\cdots;3)$，$(f(12);3;\cdots;3)$，$(f(13);3;\cdots;3)$三者之一。

**注1** 与定理 6 和定理 7 中的函数 $g(n)$，$h(n)$有关的问题容易推广到对正实数 $x$ 的所有剖分上取最大值(在这类问题中，人们偏爱于把实数 $e=2.718\cdots$作为剖分的一个“部分”)。如果定义$\binom{x}{a}$为

$$\Gamma(x+1)/\Gamma(a+1)\Gamma(x-a+1),$$

那么能得到定理 8 中 $f(n)$的一个类似推广。

**注2** 假设 $j(n)=\max \text{L.C.M.}(a_1,a_2,\cdots,a_k)$①，其中最大值也是关于 $n$ 的所有剖分取的。因为“最优化原则”，即

$$j(n)=\max_{1<a<n} \text{L.C.M.}(a,j(n-a))$$

不再成立（问题出在使 $a\cdot j(n-a)$ 达到最大的 $a$ 值可能与 $j(n-a)$ 不互素)，所以其解有一个十分不同的特征.这个最大值问题可参见文献[ 4 ]。

## 7. 简化公式

定理 1 的结果可作如下推广：

$$\binom{\binom{n}{1}}{2}=1\cdot\binom{n}{2},$$

① L.C.M表示最小公倍数。——译者注

$$\binom{\binom{n}{2}}{2}=3\binom{n+1}{4},$$

$$\binom{\binom{n}{3}}{2}=6\binom{n+2}{6}+3\binom{n+1}{6}+\binom{n}{6},$$

$$\binom{\binom{n}{4}}{2}=10\binom{n+3}{8}+15\binom{n+2}{8}+10\binom{n+1}{8},$$

$$\binom{\binom{n}{5}}{2}=15\binom{n+4}{10}+45\binom{n+3}{10}+55\binom{n+2}{10}$$

$$+10\binom{n+1}{10}+\binom{n}{10}.$$

其一般形式是

$$\binom{\binom{n}{b}}{a}=\sum_{j=1}^{b}\left[\binom{\binom{b}{j}+\varepsilon_j}{2}\right]\binom{n+b-j}{2b},$$

其中

$$\varepsilon_j=\begin{cases}1, & n\text{ 为奇数,}\\ 0, & n\text{ 为偶数.}\end{cases}$$

这个公式可解释为：因为$\binom{\binom{n}{b}}{a}$是从由 个物体构成的集合里选取 $b$-元子集对的方法数，所以我们基本上必须挑选

$2b$个物体，但它们不一定互不相同。表达式$\binom{n+b-j}{2b}$是指在原来的 $n$ 个物体的集合中添加 $b-j$ 个“百搭”后再挑选 $2b$ 个物体的方法数，其中 $0 \leqslant b-j \leqslant b-1$，这两个不同的 $b$-元子集中的元素可以互不相同，也可以至多有 $b-1$ 个元素相同。结果发现这个表达式合适的系数是

$$\left[\frac{\binom{b}{j}}{2}\right]\ (j\text{ 为偶}),\quad \left[\frac{\binom{b}{j}+1}{2}\right]\ (j\text{ 为奇}).$$

持怀疑态度的读者可能很想知道

$$\left(\!\!\binom{\binom{n}{b}}{2}\!\!\right)$$

用一些形如

$$\left[\frac{\binom{b}{j}}{2}\right]$$

的同类项来表示的简化公式的优点是什么？我们的回答是：当 $b$ 固定时，即使对于任意大的（或变化的）$n$，系数

$$\left[\frac{\binom{b}{j}}{2}\right]\text{ 和 }\left[\frac{\binom{b}{j}+1}{2}\right]$$

都是固定的且是比较小的数。这个简化公式也可写为

$$\left(\!\!\binom{\binom{n}{b}}{2}\!\!\right)=\sum_{k=0}^{b-1}\left[\frac{\binom{b}{k}+\varepsilon_k^b}{2}\right]\binom{n+k}{2b}.$$

其中

$$\varepsilon_k^b=\begin{cases}0, & \text{如果 } b\equiv k \pmod 2)\\ 1. & \text{如果 } b\not\equiv k \pmod 2).\end{cases}$$

另一简化公式是

$$\left(\begin{pmatrix} n \\ b \end{pmatrix} \atop 2\right)=\sum_{j=1}^{b}\binom{2j-1}{j}\binom{b+j}{2j}\binom{n}{b+j},$$

等式右边的项有 3 个因式，它们没有 3-层的二项式系数出现。这里不用在纸牌中添加“百搭”而只是在 $n$ 张牌中挑选 $b+j$ 张，并将其中的 $b-j$ 张用在两个 $b$-元子集里。

取决于人们的需要，$(n;2;3)$ 可给出或是“递增”或是“递减”的简化：

$$\left(\begin{pmatrix} n \\ 2 \end{pmatrix} \atop 3\right)=\binom{n+1}{6}+13\binom{n+2}{6}+\binom{n+3}{6}$$

$$=\binom{n}{3}+16\binom{n}{4}+30\binom{n}{5}+15\binom{n}{6}.$$

类似地有

$$\left(\begin{pmatrix} n \\ 3 \end{pmatrix} \atop 3\right)=4\binom{n+5}{9}+80\binom{n+4}{9}+120\binom{n+3}{9}$$

$$+65\binom{n+2}{9}+10\binom{n+1}{9}+\binom{n}{9}$$

$$=4\binom{n}{4}+100\binom{n}{5}+480\binom{n}{6}+945\binom{n}{7}$$

$$+840\binom{n}{8}+280\binom{n}{9}.$$

对最一般的情况$\left(\binom{n}{b}\atop a\right)$，目前还未发现简单的简化公式。

## 参 考 文 献


[1] S.W.Golomb, Problem No.232, Pi Mu Epsilon Journal, 5(2)(Spring)87.

[2] ——, Problem No. 223, Pi Mu Epsilon Journal, 5(1)(Fall 1969) 24 (Solution in 5(3)( Fall 1970)137).

[3] ——, Problem E 2118, *Amer.Math.monthly*, 75(1968)898.

[4] Jean-Louis Nicolas, Surl'ordre maximum d'un élément dans le groupe $S_n$ des permutations, Acta Arith., 14(1967/68), 315—332.


（刘　勇译，朱学贤校）

# 二项式型恒等式与超几何级数①

Ranjan Roy

## 1. 引言

在数学的许多领域和理论物理研究中，我们常会遇到一些组合问题。这些问题的解决依赖于求二项式系数乘积的和。有好几种方法可以用来求这类和，其中某些方法可以在 Riordan [13] 与 Kunth [11] 的书中找到。Kunth 还列举了许多二项式系数乘积的和的公式.我们简称之为二项式型恒等式.为了证明这些恒等式,很多优秀数学家曾耗费了大量的精力.事实上，恒等式并不如人们想像的那么多，从本质上来说，只有为数不多的恒等式，因此可以说大量的数学才能是白白地浪费了。许多数学家之所以不能看清某一恒等式等价于另一已知式子，是由于二项式系数的记号把他们的思路引入了歧途。记号也掩盖了各种不同形式的和在其本质上的恒同性。为此，我们将不袭用公认的二项式系数记号，而把它拆散后重新组合成其它的形式.

文中我们将指出如何把二项式系数乘积的级数化为标准形式,使级数的特征容易辨别. Euler 与 Guass 曾用过这一方法.这些标准级数就是熟知的超几何级数.超几何级数的记号

① Binomial identities and hypergeometric series, *Amer. Math. monthly*, 94(1987), 36—46.

不仅便于应用，也明显地表示和的某些重要特征，使人们能对和进行分类并将其标准化。这里我们考虑一些最重要的超越恒等式．它包含了大多数二项式系数乘积的求和。下面我们要指出，为什么超几何级数被取作标准形式，并通过几个来自不同出处的例子，说明如何将二项式系数乘积的和化为标准形式。

下节我们要定义并叙述四个主要超几何恒等式，因为它们的证明与应用没有什么联系，而我们关心的是它们的应用，所以将证明推迟到第四节。第三节是讲述例子。读者也可查阅 Andrews [2]，他的观点与本文是一致的，但例子比我们丰富。关于恒等式的有趣历史，可参看 Askey [4,5]。

## 2. 超几何级数

一级数

$$\sum c_n, \tag{2.1}$$

如果比 $c_{n+1}/c_n$ 是 $n$ 的有理函数，则称级数为超几何级数（或广义超几何级数）。例如 $c_n$ 是二项式系数的乘积，则 $c_{n+1}/c_n$ 即为 $n$ 的有理函数。对有理函数总可分解成如下形式：

$$\frac{c_{n+1}}{c_n}=\frac{(n+a_1)\cdots(n+a_p)x}{(n+b_1)\cdots(n+b_q)(n+1)}. \tag{2.2}$$

因此级数 (2.1) 可表示成下面级数

$$\sum_{n=0}^{\infty}\frac{(a_1)_n\cdots(a_p)_n}{(b_1)_n\cdots(b_q)_n}\frac{x^n}{n!} \tag{2.3}$$

乘以常数 $c_0$，这里 $(a)_n$ 定义为

$$(a)_n=a(a+1)\cdots(a+n-1),\quad (a)_0=1, \tag{2.4}$$

称为拟阶乘（shifted factorial）。我们把级数（2.3）记作：

$$ {}_pF_q\left(\begin{matrix}a_1,\cdots,a_p\\ b_1,\cdots,b_q\end{matrix};\ x\right). $$

若 $x=1$，上式书写时略去 $x$。$p=q+1$ 情形最为常见，所以我们只讨论这种情形。参数 $a_i$ 与 $b_i$ 可以是复数，但对我们的应用来说，只考虑实参数就够了。我们这里感兴趣的是有限级数，若是无穷级数，为了保证级数收敛，需要假定 $|x|<1$，或 $x=1$ 与 $(\sum b_i-\sum a_i)>0$。

应用到二项式型恒等式时，$q$ 的值通常很小，一般为 1 或 2；又参数 $a_1,\cdots,a_p,b_1,\cdots,b_q$ 满足某种关系，这关系对级数的分类和保证方法的有效性起了很重要的作用。当 $x=1$，$a_i$ 中有一为负整数（此时级数（2.3）为一有限和），并满足

$$ k+\sum a_i=\sum b_i \tag{2.5} $$

时，我们称级数为 $k$-平衡（$k$-balanced）级数。$k=1$ 情形最重要，这时级数称为平衡级数或 Saalschütz 级数。若取 $q=p-1$，且满足

$$ 1+a_1=b_1+a_2=\cdots=b_{p-1}+a_p, \tag{2.6} $$

这时级数称为良均衡（well-poised）级数。

现在我们叙述在实践中经常遇到的四个超几何级数。为数众多的二项式型恒等式都可化为所述超几何级数。从现在起，$n$ 表示正整数。

Chu-Vandermonde 恒等式：

$$ {}_2F_1\left(\begin{matrix}-n, & -b\\ & c\end{matrix}\right)=\frac{(c+b)_n}{(c)_n}. \tag{2.7} $$

Pfaff-Saalschütz 恒等式：

$$ {}_3F_2\begin{pmatrix} -n, & -a, & -b \\ & c, & d \end{pmatrix} = \frac{(c+a)_n (c+b)_n}{(c)_n (c+a+b)_n}. \quad (2.8) $$

这里 $d=1-a-b-n-c$，即级数是平衡级数。

Sheppard – Andersen 恒等式：

$$ {}_3F_2\begin{pmatrix} -n, -a, & -b, \\ c, & 2-n-a-b-c \end{pmatrix} $$

$$ = \frac{(c+b-1)_n (c+a)_n}{(c+a+b-1)_n (c)_n}\left[1+\frac{na}{(c+b-1)(a+c+n-1)}\right]. \quad (2.9) $$

级数（2.9）是2-平衡级数。

前两个恒等式发现较早，Chu① 在 1303 年公布了第一个恒等式。第二个恒等式是 Pfaff [12] 在 1797 年得到的，后来被人们遗忘，直到 1890 年由 Saalschütz [14] 再次发现该式。Sheppard [15] 在 1912年公布了 $k$-平衡级数 ${}_3F_2$ 求和公式，后来 Andersen [1] 在概率论的一项工作中也得到 $k=2$ 时的第三个恒等式。

最后一个恒等式是关于良均衡级数 ${}_3F_2$，它是属于 Dixon [7] 的。这个恒等式最好用 Gamma 函数 $\Gamma(s)$ 来表示，$\Gamma(s)$的定义为

$$ \Gamma(s) = \int_0^\infty x^{s-1} e^{-x}\, dx \, (s>0). \quad (2.10) $$

① 朱世杰是我国元代的杰出数学家，生平不详。著有《四元玉鉴》(1303年) 和《算学启蒙》(1299年)。在《四元玉鉴》一书中提出了求高阶等差级数和的"垛积术"和"招差术"，西方的 J.Gregory (1670 年) 和 Newton (1676，1678年) 对这方面的研究要晚三百六十多年。关于他的生平和工作的详细介绍可参看：钱宝琮主编的《中国数学史》，李俨，杜石然著的《中国古代数学简史》，以及杜石然的文章"朱世杰研究"(钱宝琮等著《宋元数学史论文集》，第 166—209页)。——校注

然后利用

$$\Gamma(s+1)=s\Gamma(s) \tag{2.11}$$

开拓到除 $s=0,-1,-2,\cdots$ 之外的所有实数。

Gamma 函数有下列重要性质：

$$\Gamma(s+1)=s! \quad (s \text{ 是非负整数}), \tag{2.12}$$

$$\Gamma(s)\Gamma(1-s)=\frac{\pi}{\sin s\pi} \quad (s \text{ 非整数}), \tag{2.13}$$

与

$$\lim_{s\to+\infty}\frac{\Gamma(s+1)e^{s}}{\sqrt{2\pi s}\cdot s^{s}}=1 \text{(Stirling 公式)}. \tag{2.14}$$

Dixon 的恒等式就是：

$$\begin{aligned}
&{}_3F_2\left(\begin{matrix} a, & -b, & -c \\ & a+b+1, & a+c+1\end{matrix}\right) \\
&=\frac{\Gamma\left(1+\frac{a}{2}\right)\cdot\Gamma(1+a+b)\cdot\Gamma(1+a+c)}{\Gamma(1+a)\cdot\Gamma\left(1+\frac{a}{2}+b\right)\cdot\Gamma\left(1+\frac{a}{2}+c\right)} \\
&\quad\cdot\frac{\Gamma\left(1+\frac{a}{2}+b+c\right)}{\Gamma(1+a+b+c)}.
\end{aligned} \tag{2.15}$$

注意级数可以是无穷级数，这时为了保证级数收敛，要求 $a+2b+2c+2>0$。

当化二项式系数乘积的和式为超几何级数时，我们需要下面这些初等恒等式。

$$(f)_{n-k}=\frac{(-1)^k(f)_n}{(-f-n+1)_k}\quad (n\geqslant k). \tag{2.16}$$

当$f=1$时，得

$$(n-k)!=\frac{(-1)^k n!}{(-n)_k}, \tag{2.17}$$

$$(a-n)_k=(-1)^k(-a+n-k+1)_k \tag{2.18}$$

与

$$(a)_{2k}=2^{2k}\left(\frac{a}{2}\right)_k\left(\frac{a+1}{2}\right)_k. \tag{2.19}$$

这些式子的证明非常容易。例如证(2.16)式，只要把右端按定义写出即得左端：

$$\frac{f(f+1)\cdots(f+n-1)}{(f+n-1)\cdots(f+n-k)}=f(f+1)\cdots(f+n-k-1)$$

$$=(f)_{n-k}. \tag{2.20}$$

式子(2.18)与(2.19)一样可证。

### 3. 例

我们通过几个例子来说明如何把二项式系数乘积的和化为超几何级数。进一步说明虽有许多形式各异的和，最终都可化为上述几类超几何级数。为了坚信这一点，读者可以从教科书或 Monthly (指“美国数学月刊”——译注)的问题栏中摘取一些二项式型恒等式，用这里所述的方法亲自验证一下。

例1　求下面的和：

$$S=\sum_{j=0}^{n}(-1)^j\frac{\binom{k}{j}\binom{k-1-j}{n-j}}{j+1}\quad (k\geqslant n+1). \tag{3.1}$$

这是1984年12月的 Monthly 上问题E 3065。现在我们用超几何级数工具来解这个问题。事实上它可化为 Chu-Vandermonde 级数。为此把(3.1)的项用阶乘表示，得

$$S=\sum_{j=0}^{n}(-1)^j\frac{k!}{j!(k-j)!}\cdot\frac{(k-1-j)!}{(n-j)!(k-1-n)!}\frac{1}{j+1}. \tag{3.2}$$

再把与 $j$ 无关的因子提出求和号之外，且将余下的因子用拟阶乘 $(a)_j$ 来表示，由(2.17)可得

$$S=\frac{(k-1)!}{(k-1-n)!n!}\sum_{j=0}^{n}\frac{(-k)_j}{j!}\frac{(-n)_j}{(-k+1)_j}\frac{1}{j+1}$$

$$=\binom{k-1}{n}\sum_{j=0}^{n}\frac{(-k)_j\ (-n)_j}{(1)_{j+1}\ (-k+1)_j}. \tag{3.3}$$

为了完全与超几何级数的形式一致，利用 $(a)_j=\frac{1}{a-1}\cdot(a-1)_{j+1}\ (a\neq 1)$，将上式改写为：

$$S=\binom{k-1}{n}\frac{(-k)}{(n+1)(k+1)}\sum_{j=0}^{n}\frac{(-k-1)_{j+1}(-n-1)_{j+1}}{(1)_{j+1}(-k)_{j+1}}. \tag{3.4}$$

令 $j+1=l$，上面求和式可改写成：

$$\sum_{l=1}^{n+1}\frac{(-k-1)_l(-n-1)_l}{(1)_l(-k)_l}. \tag{3.5}$$

上式加1即为超几何级数 ${}_2F_1\left(\begin{matrix}-n-1 & -k-1\\ & -k\end{matrix}\right)$. 于是由(2.7)得：

$$S=\binom{k-1}{n}\frac{k}{(n+1)(k+1)}\left[1-{}_2F_1\left(\begin{matrix}-n-1, & -k-1\\ & -k\end{matrix}\right)\right]$$

$$=\binom{k-1}{n}\frac{k}{(n+1)(k+1)}\left[1-\frac{(1)_{n+1}}{(-k)_{n+1}}\right]. \qquad (3.6)$$

经化简得到：

$$S=\frac{1}{k+1}\left[\binom{k}{n+1}+(-1)^n\right].$$

读者不难用Chu-Vandermonde 恒 等 式 证明下列二项式型恒等式：

$$\sum_{k=0}^{n}\binom{r+k}{k}=\binom{r+n+1}{n}, \qquad (3.7)$$

这里 $r$ 是实数，$n$ 为非负整数，$\binom{r}{k}=\frac{(r-k+1)_k}{k!}$（$k$为非负整数）；

$$\sum_{k=0}^{r}\binom{r}{k}\binom{s}{n+k}=\binom{r+s}{r+n} \qquad (3.8)$$

这里 $s$ 是实数，$r$ 为非负整数，$n$ 为整数，当 $n$ 为 负 整数时，定义$\binom{s}{n}=0$；

$$\sum_{k=0}^{r}(-1)^k\binom{r}{k}\binom{s+k}{n}=(-1)^r\binom{s}{n-r}, \qquad (3.9)$$

这里字母的意义同(3.8)。通过这些例子，使我们看 到 许多二项式型恒等式的证明只要按例行程序演算就行了。

例2　求下面的和：

$$S=\sum_{k\geqslant 0}\binom{n+k}{m+2k}\binom{2k}{k}\frac{(-1)^k}{k+1}. \qquad (3.10)$$

这是 Knuth[11] 书上最难的二项式型恒等式问题。为了把它化到 Pfaff-Saalschütz 恒等式，我们利用(2.17)把 $S$ 改写成

$$S=\sum_{k\geqslant 0}\frac{(n+k)!(2k)!(-1)^k}{(m+2k)!(n-m-k)!k!k!(k+1)}$$

$$=\frac{n!}{m!(n-m)!}\sum_{k\geqslant 0}\frac{(n+1)_k(-n+m)_k(1)_{2k}}{(m+1)_{2k}(1)_k(1)_k}. \qquad (3.11)$$

对 $(1)_{2k}$，$(m+1)_{2k}$ 应用二重公式(2.19)，得到

$$S=\binom{n}{m}\sum_{k\geqslant 0}\frac{(n+1)_k(-n+m)_k(1)_k\left(\frac{1}{2}\right)_k}{\left(\frac{m+1}{2}\right)_k\left(\frac{m}{2}+1\right)_k(1)_k(1)_k}$$

$$=\binom{n}{m}\frac{\frac{m}{2}\left(\frac{m-1}{2}\right)}{(-1-n+m)n\left(-\frac{1}{2}\right)}$$

$$\times\sum_{k\geqslant 0}\frac{(-1-n+m)_{k+1}(n)_{k+1}\left(-\frac{1}{2}\right)_{k+1}}{\left(\frac{m-1}{2}\right)_{k+1}\left(\frac{m}{2}\right)_{k+1}(1)_{k+1}}$$

$$=\binom{n}{m}\frac{m(m-1)}{2n(n+1-m)}$$

$$\times\left[{}_3F_2\begin{pmatrix}-1-n+m, & n\ , & -\frac{1}{2}\\ & \frac{m-1}{2}, & \frac{m}{2}\end{pmatrix}-1\right]$$

(由(2.8))

$$=\frac{(n-1)!}{2(m-2)!(n-m+1)!}$$

$$\times\left[\frac{\left(\frac{m-1}{2}-n\right)_{n+1-m}\left(\frac{m}{2}\right)_{n+1-m}}{\left(\frac{m-1}{2}\right)_{n+1-m}\left(\frac{m}{2}-n\right)_{n+1-m}}-1\right]$$

(由(2.18))

$$=\frac{(n-1)!}{2(m-2)!(n-m+1)!}$$

$$\times\left[\frac{\left(\frac{m+1}{2}\right)_{n+1-m}\left(\frac{m}{2}\right)_{n+1-m}}{\left(\frac{m-1}{2}\right)_{n+1-m}\left(\frac{m}{2}\right)_{n+1-m}}-1\right]$$

$$=\frac{(n-1)!}{2(m-2)!(n-m+1)!}\left[\frac{n+\frac{1}{2}-\frac{m}{2}}{\frac{m-1}{2}}-1\right]$$

$$=\binom{n-1}{m-1}. \tag{3.12}$$

在 Knuth [11] 书上还有一求

$$\sum_{k>0}\binom{m-r+s}{k}\binom{n+r-s}{n-k}\binom{r+k}{m+n}$$

和的问题，读者可以证明，级数同样能化到 Pfaff-Saalschütz 级数，求出它的和为 $\binom{r}{m}\binom{s}{n}$.

I. J. Good[9] 在概率论的工作中遇到求下面级数和：

$$\sum_{\nu=0}^{s}(-1)^{\nu}\binom{\beta}{\nu}\binom{\beta+s-\nu}{\beta}\frac{\alpha}{\alpha+s-\nu}. \qquad (3.13)$$

他声称当$\alpha>\beta$时，要看出和取正值是困难的。但用我们的方法，(3.13)等价于下式

$$\frac{(\beta+s)!}{s!\beta!}\frac{\alpha}{\alpha+s}{}_3F_2\left(\begin{matrix}-s, & -\beta, & -\alpha-s\\ & -\beta-s, & -\alpha-s+1\end{matrix}\right). \qquad (3.14)$$

我们再次回到 Pfaff-Saalschütz 级数(2.8)，从而求出(3.13)的和为

$$\frac{(\alpha-\beta)_s}{(\alpha+1)_s}. \qquad (3.15)$$

即可看出$\alpha>\beta$时和取正值。

现在考虑恒等式

$$\sum_{j=0}^{k}\binom{k}{j}^2\binom{n+2k-j}{2k}=\binom{n+k}{k}^2, \qquad (3.16)$$

这里$k,n$是非负整数。Takács[16] 简短地回顾了五十多年来关于这个等式各种证法的历史，有些证法是相当复杂的。L. Carlitz 指出，它也是 Pfaff-Saalschütz 级数(2.8)的一种特殊形式，读者不难证实这一点。

上面给出的级数都可化到平衡级数${}_3F_2$。而摘自 Riordan

[见文献13，p.87]书上的例子，即求下面式子

$$\sum_{k=0}^{n}\frac{2n}{n+k}\binom{n+k}{2k}\binom{2k}{k}(k+p)^{-1}(-1)^{k} \qquad (3.17)$$

的和，它只能化到2-平衡级数${}_3F_2$。事实上它可化为：

$$\frac{2}{p}\,{}_3F_2\left(\begin{matrix}-n, & n, & p\\ & 1, & p+1\end{matrix}\right). \qquad (3.18)$$

由(2.9)即可求出它的和为$-\dfrac{2n}{p(p+n)}\dfrac{(-p)_n}{(p)_n}$,

**例3** 证明等式：

$$\sum_{k=-l}^{l}(-1)^{k}\binom{2l}{l+k}\binom{2m}{m+k}\binom{2n}{n+k}$$

$$=\frac{(l+m+n)!\,(2l)!\,(2m)!\,(2n)!}{(l+m)!\,(m+n)!\,(n+l)!\,l!\,m!\,n!}, \qquad (3.19)$$

这里$l=\min(l,m,n)$。

这是Knuth[11]书内第73页上的问题62。它可化为由式(2.15)给出的Dixon的良均衡级数${}_3F_2$。令$j=k+l$，上式左端变为

$$(-1)^{l}\sum_{j=0}^{2l}(-1)^{j}\binom{2l}{j}\binom{2m}{m-l+j}\binom{2n}{n-l+j}. \qquad (3.20)$$

再用上面的推导方法，可化为如下超几何级数

$$\frac{(-1)^{l}(2m)!\,(2n)!}{(m-l)!\,(m+l)!\,(n-l)!\,(n+l)!}$$

$$\times {}_3F_2\begin{pmatrix} -2l, & -m-l, & -n-l \\ & m-l+1, & n-l+1 \end{pmatrix}. \qquad (3.21)$$

这是 Dixon 的良均衡级数，但不能直接应用 Dixon 的结果，因为 $\Gamma(1-l)/\Gamma(1-2l)$ 没有定义。为了克服这一困难，我们考虑下面的 Dixon 恒等式。

$${}_3F_2\begin{pmatrix} -2l-2\varepsilon, & -m-l-\varepsilon, & -n-l-\varepsilon \\ & m-l-\varepsilon+1, & n-l-\varepsilon+1 \end{pmatrix}$$

$$=\frac{\Gamma(1-l-\varepsilon)\Gamma(1+m-l-\varepsilon)\Gamma(1+n-l-\varepsilon)}{\Gamma(1-2l-2\varepsilon)\Gamma(1+m)\Gamma(1+n)}$$

$$\times\frac{\Gamma(1+m+n+l+\varepsilon)}{\Gamma(1+m+n)}. \qquad (3.22)$$

对 (3.22) 应用公式 (2.13)，得到

$$\frac{\sin\pi(2l+2\varepsilon)}{\sin\pi(l+\varepsilon)}\cdot\frac{\Gamma(2l+2\varepsilon)}{\Gamma(l+\varepsilon)}$$

$$\frac{\Gamma(1+m-l-\varepsilon)\Gamma(1+n-l-\varepsilon)\Gamma(1+m+n+l+\varepsilon)}{\Gamma(1+m)\Gamma(1+n)\Gamma(1+m+n)}.$$

令 $\varepsilon\to0$，可得

$${}_3F_2\begin{pmatrix} -2l, & -m-l, & -n-l \\ & m-l+1, & n-l+1 \end{pmatrix}$$

$$=(-1)^l 2\cdot\frac{\Gamma(2l)}{\Gamma(l)}$$

$$\cdot\frac{\Gamma(1+m-l)\Gamma(1+n-l)\Gamma(1+m+n+l)}{\Gamma(1+m)\Gamma(1+n)\Gamma(1+m+n)}.$$

再结合(3.21)即证明了(3.19)。

Riordan[13, p.89]所给出的级数

$$\sum_{k=1}^{m} 2k\binom{2p}{k+p}\binom{2n}{k+n} \tag{3.23}$$

也能化为 Dixon 和，这里 $m=\min(p,n)$。

## 4. 超几何恒等式证明

这节我们给出第 2 节所述恒等式的证明，当然也有其它证法，读者可参看 Askey[3] 或 Bailey[6]，其中 Askey 的分析证明很好地说明了为什么等式成立。

首先证明 Pfaff-Saalschütz 恒等式。证明之前，先利用 Gamma 函数性质(2.11)，把(2.8)式改写成：

$${}_3F_2\left(\begin{matrix}-a, & -b, & -n \\ & c, & d\end{matrix}\right)$$

$$=\frac{\Gamma(a+c+n)\Gamma(b+c+n)\Gamma(a+b+c)\Gamma(c)}{\Gamma(c+a)\Gamma(b+c)\Gamma(c+n)\Gamma(a+b+c+n)}. \tag{4.1}$$

由此看出式子关于 $a,b,n$ 是对称的。

我们采用 Dougall[8] 的方法来证明(2.8)。因为 $d=1-a-b-n-c$，为证(2.8)只需证：

$$(c)_n(c+a+b)_n\sum_{j=0}^{n}\frac{(-n)_j(-a)_j(-b)_j}{j!(c)_j(1-a-b-n-c)_j}$$

$$=(c+a)_n(c+b)_n. \tag{4.2}$$

由(2.16)可得出

$$\frac{(c+a+b)_n}{(1-a-b-n-c)_j}=(-1)^j(c+a+b)_{n-j}. \tag{4.3}$$

从而即可看出(4.2)的两端都是关于 $b$ 的 $n$ 次多项式。所以证明(4.2)只需证明它在 $b$ 的$n+1$个不同的值处等式(4.2)成立即成。当 $n=0$ 时，等式(4.2)显然是成立的。假设等式对 $n=0,1,\cdots,k-1$ 时成立，来证等式 $n=k$ 时也成立；即证等式

$$(c)_k(c+a+b)_k\sum_{j=0}^{k}\frac{(-k)_j(-a)_j(-b)_j}{j!(c)_j(1-a-b-k-c)_j}$$

$$=(c+a)_k(c+b)_k \tag{4.2}$$

成立。由于 $k$ 与 $b$ 的对称性，及 $b$ 为非负整数时上式等价于：

$$(c)_b(c+a+k)_b\sum_{j=0}^{b}\frac{(-b)_j(-a)_j(-k)_j}{j!(c)_j(1-a-k-b-c)_j}$$

$$=(c+a)_b(c+k)_b$$

再由归纳法的假设，知上式当 $b=0,1,\cdots,k-1$ 时成立，故(4.2)式当 $b=0,1,\cdots,k-1$ 时成立。我们只要再找出 $b$ 的一个值使(4.2)式成立即成。由(4.3)看出，只要取$b=-a-c$，(4.2)式两端均等于$(c+a)_k(-a)_k$。既然(4.2)在 $b$ 的 $k+1$ 个不同点处成立，因而(4.2)式对所有 $b$ 成立，(2.8)式得证。

从(4.1)出发，我们可以导出 Chu-Vandermonde 恒等式。令 $a=m$ 为正整数，并令 $n\to+\infty$，应用 Stirling 公式(2.14)得

$${}_2F_1\left(\begin{matrix}-m, & -b\\ & c\end{matrix}\right)=\frac{\Gamma(m+b+c)\Gamma(c)}{\Gamma(c+m)\Gamma(b+c)},$$

再应用(2.11)即得

$$
{}_2F_1\begin{pmatrix} -m, & -b \\ & c \end{pmatrix} = \frac{(c+b)_m}{(c)_m} \tag{4.4}
$$

如同(2.8)的证明，同样可以用数学归纳法证明 Sheppard-Andersen 恒等式。

Dixon 恒等式的证明稍难一些。Dougall 指出用证明(4.1)的方法可以证明更一般的恒等式，即证明好的良均衡2-平衡级数 ${}_7F_6$ 的和：

$$
{}_7F_6\begin{pmatrix} a, & 1+\frac{1}{2}a, & -b, & -c, & -d, & -e, & -n \\ & \frac{1}{2}a, & 1+a+b, & 1+a+c, & 1+a+d, & 1+a+e, & 1+a+n \end{pmatrix}
$$

$$
= \frac{(1+a)_n(1+a+b+c)_n(1+a+b+d)_n}{(1+a+b)_n(1+a+c)_n(1+a+d)_n} \times \frac{(1+a+c+d)_n}{(1+a+b+c+d)_n}. \tag{4.5}
$$

这里 $n$ 是正整数，且满足 $1+2a+b+c+d+e+n=0$。这关系意味着级数是2-平衡级数。好的良均衡级数中形容词“好的”是指级数有因子

$$
\frac{\left(1+\frac{a}{2}\right)_k}{\left(\frac{a}{2}\right)_k} = \frac{a+2k}{a}.
$$

读者可自己给出(4.5)的证明，或参看 Dougall 的论文，也可参看 Bailey[6] 与 Hardy[10]。然后在 (4.5) 中令$d=-\frac{1}{2}a$，并让 $n\to\infty$，应用 Stirling 公式(2.14)与(2.11)即可推出 Dixon 恒等式。

## 参考文献


[1] Erik Sparre Andersen, Two Summation Formulae for Product Sums of Binomial Coefficients, Math. Scand., 1 (1953), 261—262.

[2] G. Andrews, Applications of Basic Hypergeometric Functions, SIAM Review. 16 (1974), 441—484.

[3] R. Askey, Lecture Notes on Special Functions, unpublished.

[4] ______, A Note on the History of Series, Mathematics Research Center Tech, Rep. 1532. University of Wisconsin, Madison.

[5] ______, How Can Mathematicians and Mathematical Historians Help Each Other, to appear.

[6] W. N. Bailey, Generalized Hypergeometric Series, Cambridge University Press. Cambridge. 1935.

[7] A. C. Dixon, Summation of a Certain Series, Proc, London Math Soc., 35 (1903), 284—289.

[8] J. Dougall, On Vandermonde's Theorem and Some More General Expansions, Proc. Edinburgh Math. Soc., 25 (1907), 33—47.

[9] I. J. Good, Random Motion and Analytic Continued Fractions, Proc. Cambridge Philos. Soc, 54 (1956), 43—47.

[10] G. H. Hardy, Ramanujan. Cambridge University

Press, Cambridge. 1940.
[11] D. Knuth, The Art of Computer Programming, Vol. I, Fundamental Algorithms. Addison-Wesley, Reading, MA, 1969.
[12] J. F. Pfaff, Disquisitiones analyticae, Helmstadii, 1797.
[13] J. Riordan, Combinatorial Identities, Krieger, New York, 1979.
[14] L. Saalschütz, Eine Summationsformel, Z. Math. Phys., 35 (1890), 186—188.
[15] W. F. Sheppard, Summation of the Coefficients of Some Terminating Hypergeometric Series, Proc. London Math. Soc., (2), 10 (1912), 469—478.
[16] L. Takács. On an Identity of Shih-Chieh Chu, Acta Sci. Math. (Szeged). 34 (1973), 383—391.

(方企勤译，潘承彪校)

# 几何平均、对数平均及算术平均不等式[①]

F. Burk

设 $0<a<b$。由凸性，对于积分 $\int_{\ln a}^{\ln b} e^x dx$ 用中点法近似及梯形法近似(见图 1)得

$$\left(e^{\frac{\ln a+\ln b}{2}}\right)(\ln b-\ln a)<\int_{\ln a}^{\ln b} e^x dx<\frac{e^{\ln b}+e^{\ln a}}{2}(\ln b-\ln a),$$

即

$$\sqrt{ab}<\frac{b-a}{\ln b-\ln a}<\frac{a+b}{2},$$

称为几何平均、对数平均及算术平均不等式。

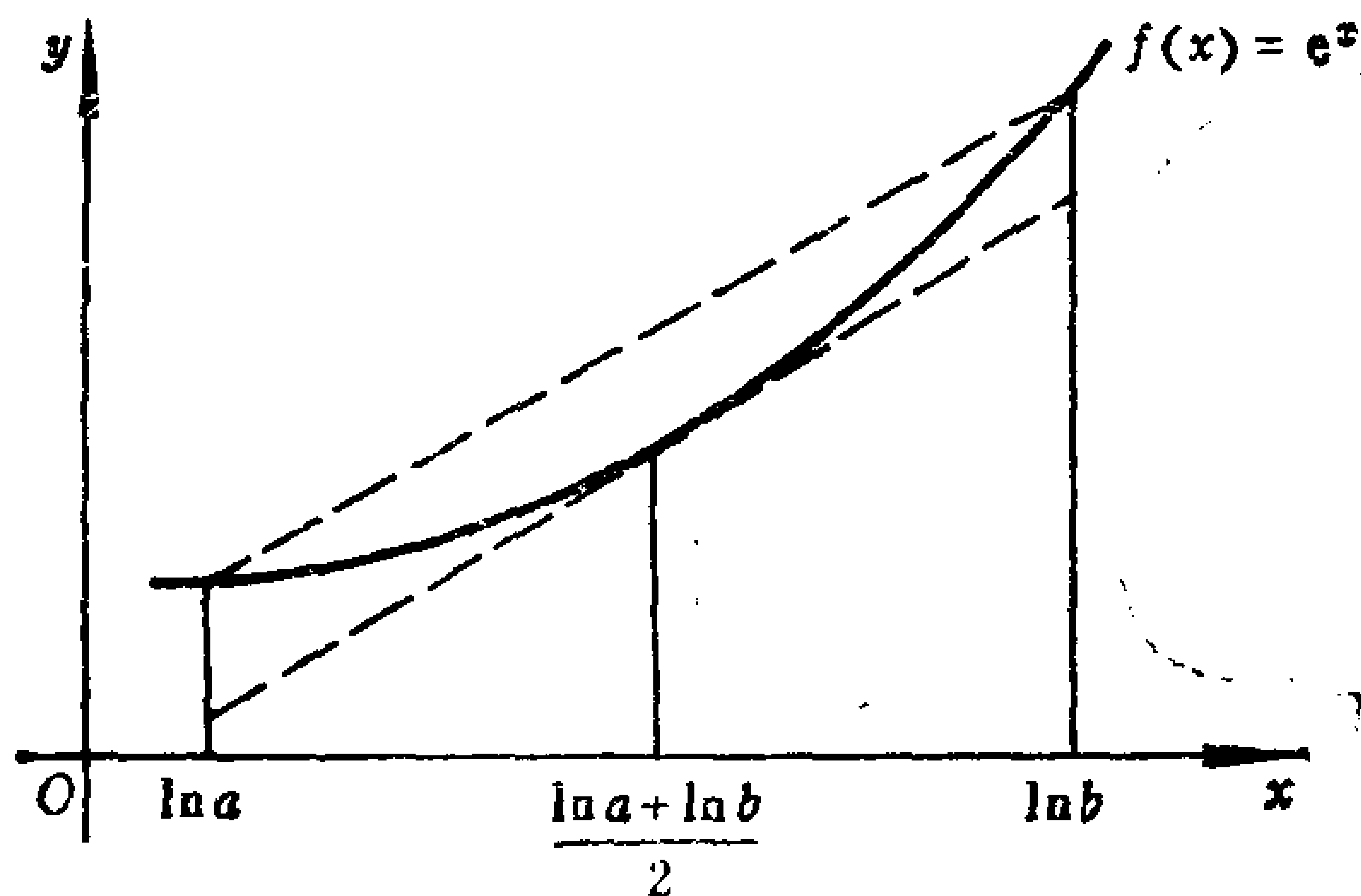



图 1

① The Geometric, Logarithmic, and Arithmetic Mean Inequality, *Amer. Math. Monthly*, 94 (1987), 527—528.

顺便说一句，Tung-Po Lin [2]证明了

$$\sqrt{ab}<\frac{b-a}{\ln b-\ln a}<\left(\frac{a^{1/3}+b^{1/3}}{2}\right)^3,$$

而且，在幂平均

$$M_p=\left(\frac{a^p+b^p}{2}\right)^{\frac{1}{p}},\qquad p\neq 0\text{ 及 }M_0=\sqrt{ab}$$

的意义下，这些估计可能是最好的。本篇短文的主要意思是表明，我们看到第2个不等式可以由应用Simpson的$\frac{3}{8}$法则(见文献[1])，即

$$\int_c^d f(x)\,\mathrm{d}x$$

$$=\left[\frac{f(c)+3f\left(\frac{2c+d}{3}\right)+3f\left(\frac{c+2d}{3}\right)+f(d)}{8}\right](d-c)$$

$$-\frac{(d-c)^5}{6480}f^{(4)}(\eta),$$

对于$c$和$d$之间的某个$\eta$成立，于$e^x$而得到，只要将其中的$c$和$d$分别换成$\ln a$和$\ln b$。我们有

$$b-a=\int_{\ln a}^{\ln b}e^x\mathrm{d}x$$

$$<\frac{e^{\ln a}+3e^{\left(\frac{2\ln a+\ln b}{3}\right)}+3e^{\left(\frac{\ln a+2\ln b}{3}\right)}+e^{\ln b}}{8}(\ln b-\ln a)$$

$$=\left(\frac{a^{1/3}+b^{1/3}}{2}\right)^3(\ln b-\ln a),$$

其中的不等式成立是由于略去了误差项。

## 参 考 文 献


[1] R.L. Burden, J.D. Faires, and A.C.Reynolds, *Numerical Analysis, second ed., Prindle, Weber,* & Schmidt, 1981.

[2] Tung-Po Lin, The power mean and the logarithmic mean, *Amer. Math. Monthly,* 81 (1974), 879—883.


（朱学贤译）

# 表整数为奇合数之和①

A.M. Vaidya

最近，Just 和Schaumberger[1]发现偶数38有一个有趣的性质。他们证明了38是不能表成为两个奇合数之和的最大偶数。在本文，我们将从两个不同的方向推广这一结果。

**定理1** 对于每个正整数 $t$，整数 $9t+20$ 不能表成 $t$ 个奇合数的和。

**证明** 对 $t=1,2$ 和 3，定理是容易证明的，因此，应用归纳法，假设定理对 $t=k-1$ 是成立的，其中 $k\geq 4$。

如果 $9k+20$ 是 $k$ 个奇合数之和，那么可以断言，这 $k$ 个奇合数中至少有一个是9，否则(由于9和15是两个最小的奇合数)这 $k$ 个数都将 $\geq 15$，于是它们的和将 $\geq 15k$。这就是说，$9k+20\geq 15k$，但这是不可能的，因为这与 $k\geq 4$ 矛盾。

因此，如果 $9k+20$ 是 $k$ 个奇合数之和，那么这 $k$ 个奇合数中必定有一个是9。从其中去掉为9的这一项，那么，就有 $9k+20-9=9(k-1)+20$，它是 $k-1$ 个奇合数之和，这与假设 $t=k-1$ 时结论成立矛盾。这就证明了定理。

**定理2** 设 $t\geq 2$ 是一个整数，如果整数 $n>9t+20$ 与 $t$ 有相同的奇偶性，那么 $n$ 可以表为 $t$ 个奇合数之和。

---

① On representing integers as sums of odd composite integers, *Math. Mag.* 48 (1975), 221—223.

**证明** 因为$n>9t+20$，并且它与$t$有相同的奇偶性，因此必有

$$n=9t+20+2m,$$

对某正整数$m$成立。

由Just与Schaumberger的结果知，$2m+38$是两个奇合数之和①，比如说它是$u$与$v$之和，这里$u$与$v$是两个奇合数。于是

$$n=9t+20+2m=9(t-2)+2m+38$$
$$=9+9+9+\cdots+9+u+v,$$

其中和数9的个数是$t-2$。因此$n$是$t$个奇合数之和。

结合定理1和定理2，就得到下面的Just和Schaumberger的结果的推广：

**定理3** 设$t\geqslant2$是一个整数，那么在所有与$t$具有相同奇偶性的整数当中，$9t+20$是不能表为$t$个奇合数之和的最大者。

如果考查Just和Schaumberger的证明，就会发现，他们实际上证明了更强的结论：每一个大于38的偶数是两个奇合数之和，其中一个是3的倍数，而另一个是5的倍数. 这样就提出了下面的问题：假设给定两个相异的奇素数$p$和$q$，那么不能表为这样两个奇合数之和的最大偶数是多少？这两个奇合数分别是$p$的倍数和$q$的倍数.我们来回答这个问题.

下面的引理是众所周知的(参看文献[2]或[3])：

**引理1** 如果$p$和$q$是互素的正整数，那么不能表为形如$pr+qs$的最大整数是$pq-p-q$。其中$r$和$s$是非负整数②。

---

① 由下面的定理4可推出这一结论。——校注

② 见华罗庚《数论导引》第一章§8定理3。——校注

这个引理使得我们能够证明下面的关于 Just 和 Schaumberger 的结果的推广：

**定理4** 如果 $p$ 和 $q$ 是两个不同的奇素数，那么不能表为这样两个奇合数之和的最大偶数是 $2pq+p+q$，其中一个是 $p$ 的倍数，另一个是 $q$ 的倍数。

**证明** 显然一个偶数 $2n$ 有一个这样的确定形式的表达式，当且仅当存在整数 $r\geqslant 0$ 和 $s\geqslant 0$，使得

$$p(2r+3)+q(2s+3)=2n.$$

也就是，当且仅当存在整数 $r\geqslant 0$ 和 $s\geqslant 0$，使得 $pr+qs=n-\frac{3}{2}(p+q)$。因此，由上面的引理，不能表成形为 $p(2r+3)+q(2s+3)$ 的最大偶数 $2n$ 对应于

$$n-\frac{3}{2}(p+q)=pq-p-q,$$

即是

$$2n=2pq+p+q.$$

**注** 令 $p=3$，$q=5$，就得到 Just 和 Schaumberger 的结果。

Shah[4] 对引理 1 有下面的推广：

**引理2** 设 $p$ 和 $q$ 是互素的正整数，$k$ 是一个确定的正整数，那么不能用 $k$ 种方法表为 $pr+qs\,(r\geqslant 0, s\geqslant 0)$ 形式的最大整数是 $kpq-p-q$①。

现在，我们能用与定理 4 相同的方法来证明 Just 和 Schaumberger 结果的如下推广形式：

---

① 利用华罗庚《数论导引》第一章 § 8 定理 3，及接着的习题 1 就不难证明这引理。——校注

**定理5** 设$p$和$q$是两个不同的奇素数。那么不能用$k$种方法表为两个分别为$p$和$q$的倍数的奇合数之和的最大偶数是

$$2kpq+p+q.$$

我们称一个数写为两个奇合数之和的表示形式为其标准表示式。对一个正整数$k$，所有不能写成为$k$种标准表示式中的最大整数，我们记之为$F(k)$，很显然，对$F(k)$我们有

$$F(1)=38,\quad F(2)=F(3)=68,\quad F(4)=94,$$
$$F(5)=122,\quad F(6)=F(7)=128,\quad F(8)=136.$$

## 参 考 文 献


[1] Erwin Just and Schaumberger, A curious property of the integer 38, *Math. Magazine*, 46 (1973) , 221.

[2] W. Leveque, Topics in Number Theory, Vol.I. Addison-Wesley, Reading. 1956. p.22.

[3] Problem E 1637, Amer. Math. Monthly, 70(1963), 1005, and 71 (1964),799.

[4] A.P. Shah, Contribution to…a problem of Frobenius, Ph. D. thesis submitted to Gujarat University, 1970.


（李　伟译，潘承彪校）

# 条件极值的二阶导数判别法①

D. Spring

## 1. 引言

在许许多多的数学及应用的课题中出现的一个著名的数学问题是求函数的极值(极大值或极小值)，其中的函数满足一些附加的约束条件. 设$f, g_a$: $U \to \boldsymbol{R}$是一些可微函数，$1 \leqslant a \leqslant m$，$U$是$\boldsymbol{R}^n$中的一个开集，$m < n$. 约束条件由方程$g_a = 0$，$1 \leqslant a \leqslant m$定义. 假定$A \subset U$是约束条件的解集. 问题是求函数$\tilde{f}$: $A \to \boldsymbol{R}$的极值，$\tilde{f}$是$f$在解集$A$上的限制. 举例说来，设$(x_1, \cdots, x_n)$是$\boldsymbol{R}^n$中的坐标，考虑一些简单的约束条件$x_a = 0$，$1 \leqslant a \leqslant m$. 这时解集$A = U \cap V$，$V$是$\boldsymbol{R}^n$的一个线性子空间，其中每个点的前$m$个坐标都是0. 令$\tilde{f}(x_{m+1}, \cdots, x_n) = f(0, \cdots, 0, x_{m+1}, \cdots, x_n)$，于是我们发现相对于约束条件$f$在点$a = (0, \cdots, 0, a_{m+1}, \cdots, a_n)$处取到极值当且仅当$\tilde{f}$在点$\tilde{a} = (a_{m+1}, \cdots, a_n)$处取到(通常意义下的)极值. 特别地，$\tilde{a}$是函数$\tilde{f}$的驻点，即$\tilde{f}$的所有一阶偏导数在$\tilde{a}$处取值为0. 由于在一般的约束条件时$\tilde{f}$仅仅是隐式的，所以它的极值不可能直接计算出来.

解条件极值问题的经典方法涉及到大家熟知的

① On The Second Derivative Test For Constrained Local Extrema, *Amer. Math. Monthly*, 92 (1985), 631—643.

Lagrange 乘子法(见文献[ 4 ],[ 6 ])。设 $L$: $\boldsymbol{R}^m \times U \to \boldsymbol{R}$ 是 $(m+n)$个变元的 Lagrange 函数

$$L(\lambda_1, \cdots, \lambda_m, x_1, \cdots, x_n)$$

$$= f(x_1, \cdots, x_n) + \sum_{a=1}^{m} \lambda_a g_a(x_1, \cdots, x_m),$$

固定 $a \in A$ 并出于技术上的原因假定映射 $g = (g_1, \cdots, g_m)$: $U \to \boldsymbol{R}^m$ 是 $C^1$ 类的且在 $a$ 点有最大的秩$(=m)$。这一假定保证了解集 $A$ 在 $a$ 点附近是 $\boldsymbol{R}^n$ 中的一个$(n-m)$维子流形。主要的结论是(见文献[ 4 ],[11]):相对于约束条件,如果 $f$ 在 $a \in A$处取到极值,则对于某个向量 $c = (c_1, c_2, \cdots, c_m) \in \boldsymbol{R}_m$,向量$v = (c, a) \in \boldsymbol{R}^m \times U$是 Lagrange 函数 $L$ 的一个驻点. $L$ 在 $a$ 点的所有一阶偏导数值都是 0,即

$$\frac{\partial f}{\partial x_j}(a) + \sum_{a=1}^{m} c_a \frac{\partial g_a}{\partial x_j}(a) = 0, \qquad (1.1)$$

$$1 \leqslant j \leqslant m; \ g_a(a) = 0, \ 1 \leqslant a \leqslant m.$$

因而,一般说来,条件极值问题的解一定在相应的 Lagrange 函数的驻点中找到。

但是,对于条件极值问题,驻点条件(1.1)仅仅是必要的但不充分。本文论述的中心问题就是去寻求一个合适的充分条件。

作为参考,在此叙述一下探讨充分条件的传统方法是有益的。在无条件极值问题时,$m=0$,$L=f$。设 $a \in U$ 是 $f$ 的驻点:$\frac{\partial f}{\partial x_i}(a) = 0$,$1 \leqslant i \leqslant n$。假定 $f$ 是 $C^3$ 类的,在 $a$ 点展开 $f$ 成 Taylor 级数($h = (h_1, \cdots, h_n) \in \boldsymbol{R}^n$):

$$f(a+h) = f(a) + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(a) h_i h_j + R_2(h), \quad (1.2)$$

其中 $\lim\limits_{h\to 0}\frac{R_2(h)}{\|h\|^2}=0$。驻点 $a$ 的性质是由检查在(1.2)式中出现的二次型：

$$q(h)=\frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2 f}{\partial x_i\partial x_j}(a)h_ih_j$$

来分析的。注意到 $q$ 的相伴矩阵是 Hesse 矩阵

$$H(f)(a)=\left[\frac{\partial^2 f}{\partial x_i\partial x_j}(a)\right]_{1\leq i,j\leq n}.$$

因为余项是"可以忽略的"，所以我们有下面的结论(Edwards[6,p.138])：

**命题1.1** 如果 $q$ 是正(负)定的，则 $f(a)$ 是严格的极小(大)值；如果 $q$ 既可以取到正值又可以取到负值，则 $a\in U$ 不是 $f$ 的极值点。

**注释1.1** 其余情形，即 $q$ 是退化的 ($\det Hf(a)=0$) 或者是半正定或半负定的情形，则在二阶偏导数的水平上，驻点 $a$ 的性质必定是不确定的（一个完全的讨论见§4的结论Ⅳ）。例如，设 $f(x,y)=x^2+y^4$，$g(x,y)=x^2-y^2$，则 (0,0) 点是 $f$ 和 $g$ 的唯一驻点，相应的二次型都是 $q(x,y)=x^2$，它是退化的且是半正定的，但是 $f$ 在 (0,0) 点取到严格的极小值而 $g$ 在 (0,0) 点却不取极值。

为了能够得到区分命题 1.1 中的 3 种情形的数值算法，我们观察由 Hesse 矩阵 $Hf(a)$ 的 $k\times k$ 阶顺序主子式组成的序列 $\{\Gamma_k\}_{1\leq k\leq n}=(\Gamma_1,\cdots,\Gamma_n)$ (Edwards[6,p.149])。

在 $\det Hf(a)\neq 0$ 的假定下（从而 $q$ 非退化），如果全是正号，即 +，+，+，…，则 $q$ 是正定的；如果负正号交错，即

$-,+,-,+,\cdots$，则 $q$ 是负定的；任意其它的符号序列，或者序列中有 0 项，都意味着 $q$ 既可以取到正值又可以取到负值。

在条件极值的情形，前面我们曾说过（技术性假设），$g_1=g_2=\cdots=g_m=0$ 的解集是 $\boldsymbol{R}^n$ 在 $a$ 点附近的一个子流形。设 $q_L$ 是辅助二次型

$$q_L(h)=\frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2 L}{\partial x_i \partial x_j}(v)h_i h_j,$$

相对于约束条件 $g_\alpha=0$，$1\leqslant\alpha\leqslant m$，$f$ 在 $a$ 点附近的性质是通过观察限制在 $\boldsymbol{R}^n$ 的一个线性子空间上的二次型 $q_L$ 来分析的，这个线性子空间由流形 $A$ 在 $a$ 点的切向量构成(Edwards [6,p.154])。因此，用分析的术语来说，就是相对于 $m$ 个线性约束条件

$$\sum_{i=1}^{n}\frac{\partial g_\alpha}{\partial x_i}(a)h_i=0,\quad 1\leqslant\alpha\leqslant m \tag{1.3}$$

去研究二次型 $q_L$。

**命题1.2** (Edwards[6,p.154]) 对于线性约束条件(1.3)，如果 $q_L$ 分别是正定、负定或不定的，则对于约束条件 $g_\alpha=0,1\leqslant\alpha\leqslant m$，函数 $f$ 在 $a$ 点分别取到严格极小值、严格极大值或不取极值。

**注释1.2** 命题 1.1 是命题 1.2 在 $m=0(L=f;\ \ q_L=q)$ 时的特殊情形。

**注释1.3** 其余情形，即相对于线性约束条件 (1.3)，$q_L$ 是退化的或是半正定或半负定的，则驻点 $v$ 的性质，在二阶导数的水平上，必定是不确定的（见 § 4 的 结论Ⅳ）。

为了完成命题 1.2 的分析，需要一种能把服从线性约束

条件的二次型进行分类的数值算法。H.B.Mann[10] 区分了服从线性约束条件的正定的及负定的二次型。著名的数学经济学家 G.Debreu[5] 推广了 Mann 的工作，区分了服从线性约束条件的半正定的及半负定的二次型。因为二次型既取正值又取负值的充分必要条件是既不半正定也不半负定，所以 Debreu 的算法原则上满足了区分命题 1.2 中的 3 种情形的需要。我们对于文献中有关充分条件的回顾就到这里。

本文中，我们提供另外一条途径，去研究条件极值问题的充分条件。我们的想法是在点 $a\in U$ 的邻域中作坐标变换，使解流形在 $a$ 点附近变"平坦"些，且使约束函数在新的坐标系中成为前面讨论过的一些简单约束函数 $g_\alpha(x_1,\cdots,x_n)=x_\alpha$，$1\leqslant\alpha\leqslant m$。在简单约束条件时，可以用无条件极值时区分二次型的算法得到恰当的充分条件。

设 $v=(c,a)$ 是 Lagrange 函数 $L$ 的一个驻点。我们做法的特色是，用驻点 $v$ 处算得的函数 $L$ 的 $(m+n)\times(m+n)$ 阶 Hesse 矩阵 $HL(v)$（$L$ 在 $v$ 处的二阶偏导数值构成的矩阵）叙述充分条件。具体说来就是：

$$HL(v)=\begin{bmatrix} 0 & \cdots & 0 & \dfrac{\partial g_1}{\partial x_1}(a) & \cdots & \dfrac{\partial g_1}{\partial x_n}(a) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \dfrac{\partial g_m}{\partial x_1}(a) & \cdots & \dfrac{\partial g_m}{\partial x_n}(a) \\ \dfrac{\partial g_1}{\partial x_1}(a) & \cdots & \dfrac{\partial g_m}{\partial x_1}(a) & \dfrac{\partial^2 L}{\partial x_1^2}(v) & \cdots & \dfrac{\partial^2 L}{\partial x_1\partial x_n}(v) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \dfrac{\partial g_1}{\partial x_n}(a) & \cdots & \dfrac{\partial g_m}{\partial x_n}(a) & \dfrac{\partial^2 L}{\partial x_n\partial x_1}(v) & \cdots & \dfrac{\partial^2 L}{\partial x_n^2}(v) \end{bmatrix}$$

注意到$HL(v)$的右下$n\times n$阶子矩阵是命题1.2中用到的辅助二次型

$$q_L(h)=\frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2 L}{\partial x_i\partial x_j}(v)h_ih_j$$

的Hesse矩阵。自然的问题是明确地叙述出用$HL(v)$区分下面3种重要情形的二阶导数判别法（即数值算法），这3种情形是：

相对于约束条件$g_a=0, 1\leqslant a\leqslant m$，$f$在$a$点取（严格）极小值，取（严格）极大值或不取极值（后者称为鞍状情形）。

(1.4)

定理1是主要的结果，用于区分（1.4）中的3种情形，而且在$\det HL(v)=0$时提供一种看来是新的完整的鞍状情形的判则。定理1中用到$HL(v)$的主子式序列，这一序列隐含在这一问题的某些较早的分析中（例如Hancock[9]的p.74），但可能是H.B.Mann[10]在本杂志中第一个明确引进。遗憾的是，Mann的判别法及G.Debreu的推广[5]在数学文献中几乎被遗忘了。近些年来，Mann的判别法显著地出现在将数学应用于经济学的一些教科书中，有时被称为增广的Hesse矩阵法（Ostrosky和Koch[13]，Y.Murata[12]，还可以参阅J.Marsden和A.J.Tromba[11]）。基于定理1，关键是可以看到Mann的判别法在前面讨论过的简单约束条件情形时很容易被理解。

有关Lagrange乘子的基本方法的文献是大量的，但有关二阶导数判别法的文献却很少，我意识到，在所有这些文献中，(1.4)中的鞍状情形都被忽视了(Hancock[8],[9]，Carathéodory[3]，J.Marsden和A.J.Tromba[11]，Y.Murata[12]，

F. Bowman 和 F. A. Gerard [2], K. G. Binsmore [1])。某些高等微积分的教科书给出了条件极值的充分条件但没有给出一种数值算法去区分（1.4）中的3种情形（C. H. Edwards [6], R. Courant 和 F. John [4]）。

本文试图对于条件极值问题提供一种形式明确的二阶导数判别法。从而需要更详细地研究二次型的分类。一般文献中的材料不够，为此我们在文章后面增加了一个论述二次型的附录。

## 2. 定理的叙述

设 $(y_j)_{1\leq j\leq r}$ 是非平凡的实数列(即不是所有的项全是0)。设 $y_k$ 是此序列中最后一个非零项。

**定义** （1）如果 $y_j>0$，$j=1,\cdots,k$，则称序列 $(y_j)$ 是**半正定序列**。

（2）如果$(-1)^j y_j>0, j=1,\cdots,k$，则称序列 $(y_j)$ 是**半负定序列**。

（3）如果(1)及(2)都不成立，则称非平凡序列$(y_j)$是**鞍型序列**，即，或者是某些 $y_j=0$，$j<k$，或者是符号序列不是(1)及(2)。

（4）序列$(y_j)$称为是**正(负)定序列**，如果它是半正(负)定的且 $k=r$（即最后一项 $y_r\neq0$）。

在后面的命题2.1中我们将解释这些定义。

**记号** (a) $n\times n$ 阶方阵$M$的 $j$ 阶主子式是$M$ 的 $j\times j$ 阶主子矩阵（即由$M$去掉（$n-j$）行及相应的（$n-j$）列得到的子矩阵）的行列式。

(b) 设$M$是 $n\times n$ 阶方阵。对于$\{1,\cdots,n\}$ 的每一个置换

$\pi$，记 $M(\pi)$ 是由 $\pi$ 置换 $M$ 的行及列得到的矩阵。与 $\pi$ 相连的是坐标的线性变换：$(x_1,\cdots,x_n)\to(x_{\pi(1)},\cdots,x_{\pi(n)})$。因此 $M(\pi)=E^{-1}\times M\times E$，其中 $E$ 是（依赖于 $\pi$ 的）某个正交矩阵。

本文中所需的代数方面的主要背景材料是：

**命题2.1** 设 $Q$ 是 $\boldsymbol{R}^r$ 中的实值二次型，其相伴矩阵是 $M$（在 $\boldsymbol{R}^r$ 的某组基中）。

(a) $Q$ 是正定（负定）的，当且仅当 $M$ 的顺序主子式序列 $(y_j)_{1\leqslant j\leqslant r}$ 是正定（负定）的。

(b) $Q$ 是不定的，当且仅当存在 $\{1,\cdots,r\}$ 的一个置换 $\pi$ 使得 $M(\pi)$ 的顺序主子式序列 $(y_j(\pi))_{1\leqslant j\leqslant r}$ 是非平凡的且是鞍型的。

**证明** 结论(a)是熟知的(Edwards[6,p.149])。结论(b)在文献中看来没有出现。(b)中"当"的部分是显然的且是下面引理 2.1 的一个推论（注意 $M(\pi)$ 是由 $\pi$ 置换 $\boldsymbol{R}^n$ 的给定基得到的 $Q$ 的矩阵）。(b) 中"仅当"的部分在附录中证明。

**引理2.1** 设 $q$ 是 $\boldsymbol{R}^n$ 中的实值二次型，其相伴矩阵是 $N$（在 $\boldsymbol{R}^n$ 的某组基中）.如果 $N$ 的顺序主子式序列 $(y_j)_{1\leqslant j\leqslant n}$ 是非平凡的且是鞍型的，则 $q$ 是不定的。

**证明** 设 $y_k$ 是序列中的最后一个非零项，并设 $V$ 是由给定的 $\boldsymbol{R}^n$ 的一组基的前 $k$ 个向量生成的 $\boldsymbol{R}^n$ 的子空间。用 $\tilde{q}$ 表示由限制：$\tilde{q}(x)=q(x)$，$x\in V$，得到的 $V$ 上的二次型。因为 $y_k\neq0$，所以 $\tilde{q}$ 是非退化的。又因为序列 $(y_j)_{1\leqslant j\leqslant k}$ 是鞍型的，所以由非退化的二次型的分类(Edwards [6,p.149]) 推得 $\tilde{q}$（从而 $q$）是不定的。

作了以上准备，现在叙述我们的二阶导数判别法如下：

设 $\Gamma_k=(-1)^m\times$ Hesse 矩阵 $HL(v)$ 的左上 $k$ 阶主子式，$1\leqslant k\leqslant m+n$。特别地，有 $\Gamma_{n+m}=(-1)^m\det HL(v)$。注意到 $m=0$（即无条件）时有 $L=f$ 且 $v=a\in U$ 是 $f$ 的驻点。这时，$\Gamma_k$ 是 Hesse 矩阵 $Hf(a)$ 的左上 $k$ 阶主子式，$1\leqslant k\leqslant n$。设 $\pi$ 是 $\{1,\cdots,m+n\}$ 的一个置换，则 $\Gamma_k(\pi)=(-1)^m\times$ 矩阵 $HL(v)(\pi)$ 的左上 $k$ 阶主子式，$1\leqslant k\leqslant m+n$。在判别法中，我们只用到序列

$$(\Gamma_{2m+p})_{1\leqslant p\leqslant n-m}=(\Gamma_{2m+1},\cdots,\Gamma_{2m+n})$$

及

$$(\Gamma_{2m+p}(\pi))_{1\leqslant p\leqslant n-m}.$$

**定理1**（二阶导数判别法） 设 $v=(c,a)\in \boldsymbol{R}^m\times U$ 是 Lagrange 函数 $L$ 的一个驻点，$f,g_\alpha$: $U\to\boldsymbol{R}$，$1\leqslant \alpha\leqslant m$，$m<n$，是 $C^3$ 类的可微函数及 $g=(g_1,\cdots,g_m)$：$U\to\boldsymbol{R}$ 在 $a$ 点有最大秩（$=m$）。特别地，假定

$$\frac{\partial(g_1,\cdots,g_m)}{\partial(x_1,\cdots,x_m)}(a)\neq 0$$

（如果有必要的话，将变量 $x_1,\cdots,x_n$ 重新排列）。

(a) 如果 $\Gamma_{m+n}\neq 0$（即 $\det HL(v)\neq 0$）。

(i) 如果序列 $(\Gamma_{2m+p})_{1\leqslant p\leqslant n-m}$ 是正定的，则对于约束条件，$f$ 在 $a\in U$ 处取到严格极小值。

(ii) 如果序列 $(\Gamma_{2m+p})_{1\leqslant p\leqslant n-m}$ 是负定的，则对于约束条件，$f$ 在 $a\in U$ 处取到严格极大值。

(b) 如果序列 $(\Gamma_{2m+p})_{1\leqslant p\leqslant n-m}$ 非平凡且是鞍型的（注意，我们没有假定 $\det HL(v)\neq 0$，仅仅假定对于某个 $p$，有 $\Gamma_{2m+p}\neq 0$）；则对于约束条件，$f$ 在 $a\in U$ 既不取极大值也不取极小值。我们称之为鞍状情形。

(b)′(概括的鞍状情形)。 如果存在 $HL(v)$ 中最后 $(n-m)$ 行及列的置换 $\pi$（这相应于 $\boldsymbol{R}^n$ 中最后 $(n-m)$ 个变元的置换：$(x_1,\cdots,x_n)\to(x_1,\cdots,x_m,x_{\pi(m+1)},\cdots,x_{\pi(n)})$，使得由矩阵 $HL(v)(\pi)$ 得到的序列 $(\Gamma_{2m+p}(\pi))_{1\leq p\leq n-m}$ 非平凡且是鞍型的，则对于约束条件，$f$ 在 $a\in U$ 既不取极大值也不取极小值。

**注释**

(1) 在(b)中，如果 $\Gamma_{m+n}\neq 0$（即 $\det HL(v)\neq 0$），则称此情形为非退化鞍状情形，它相当于：不是((i)或(ii))。因此，如果 $\Gamma_{m+n}\neq 0$,则二阶导数判别法是完全的。另外，如果 $\det HL(v)\neq 0$，则(b)和(b)′等价。

(2) 显然，(b)是(b)′的特殊情形(令 $\pi=\mathrm{id}$)。事实上，如果(a)不成立，则(b)一般能应用。后面的例 3 用于说明如果 $\det HL(v)=0$,则(b)不成立时可以应用(b)′.

(3) 仅有的不能确定的情形，是指 $\Gamma_{m+n}=\det HL(v)=0$ 且(b)′也不成立的情形，这时,定理 1 意指关于驻点 $v$ 的性质没有结论。按照 § 1 中命题 1.2 提供的分析充分条件的术语，这相当于说，对于线性约束条件(1.3),辅助二次型是退化的且是半正定或半负定的(完整的细节见 § 4 的结论 Ⅲ 和 Ⅳ)。在 § 4 中我们将说明，这种退化的半定情形，在二阶偏导数的水平上必定是不能确定的．因此在此意义下，定理 1 的叙述不可能再改进。

(4) 当 $m=0$（即无约束条件)且 $\Gamma_n=\det Hf(a)\neq 0$ 时，定理 1 就是大家熟知的函数在驻点处是否取极值的二阶导数判别法(Edwards[6,p.145])。要注意的是，在 $\det Hf(a)=0$

时，结论(b)和(b)′改进且完全了这一传统的判别法。

(5) 当 $n=3$ 及 $1\leqslant m\leqslant 2$ 时，定理 1 中的(a)的证明在许多教科书上都有（例如，Bowman 和 Gerard [2]）。但是，这些书上的讨论或者无视鞍状情形的存在，或者假定 $\det HL(v)\neq 0$。

(6) 当 $m=n-1$ 时，序列只有一项，即 $\Gamma_{m+n}=(-1)^m\cdot\det HL(v)$. 这时，只要假定 $g$：$U\to \boldsymbol{R}^m$ 在 $a\in U$ 处的秩等于 $m$ 就足够了（即变量不需要重新排列）。还应该看到，这时没有鞍形选择。

下面，我们用一些例子来解释定理 1 。

**例 1** $f(x,y,z)=xyz$，约束条件是 $x^2+y^2+z^2-1=0$。这个例子取自 Marsden 和 Fromba([11],p.229)。函数

$$L=xyz+\lambda(x^2+y^2+z^2-1)$$

的一个驻点是 $v=(0,1,0,0)$。在点 $v$ 处计算得 $\Gamma_3=0,\Gamma_4=-\det HL(v)=-4$。由定理 1，我们看到这是鞍状情形。但是由于没有鞍状情形，所以 Marsden 和 Tromba （不加判别地）断言判别法对 $v$ 无效，即 $v$ 的性质不能确定。

**例 2** 为了求圆 $x^2+y^2=2$ 上的点 $(x_1,x_3)$ 和直线 $x+y=4$ 上的点 $(x_2,x_4)$ 之间的最短距离（见 Edwards[6,p.156]），我们求函数

$$f(x_1,x_2,x_3,x_4)=(x_1-x_2)^2+(x_3-x_4)^2$$

在两个约束条件

$$g_1(x_1,x_2,x_3,x_4)=x_1^2+x_3^2-2=0,$$
$$g_2(x_1,x_2,x_3,x_4)=x_2+x_4-4=0,$$

下的极小值．此时 Lagrange 函数 $L=f+\lambda_1 g_1+\lambda_2 g_2$ 的驻点是：$v_1=(1,-2,1,2,1,2)$ 及 $v_2=(-3,-6,-1,2,-1,2)$．由增广的 Hesse 矩阵

$$HL(v)=\begin{bmatrix} 0 & 0 & 2x_1 & 0 & 2x_3 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 2x_1 & 0 & 2+2\lambda_1 & -2 & 0 & 0 \\ 0 & 1 & -2 & 2 & 0 & 0 \\ 2x_3 & 0 & 0 & 0 & 2+2\lambda_1 & -2 \\ 0 & 1 & 0 & 0 & -2 & 2 \end{bmatrix}$$

算得：在 $v_1$ 处 $\Gamma_5=32$，$\Gamma_6=64$；在 $v_2$ 处 $\Gamma_5=-32$，$\Gamma_6=-192$．由定理 1，可以断言，在点 $(1,2,1,2)$，即相应于圆上的 $(1,1)$ 点和直线上的 $(2,2)$ 点，$f$ 取极小值（实际上是最小值），而在点 $(-1,2,-1,2)$，即相应于圆上的 $(-1,-1)$ 点及直线上的 $(2,2)$ 点，$f$ 不取极值．这个结论在几何上是显然的．

**例 3**　$f(x,y,z,t)=y^2+z^2-t^2-x^2$，约束条件是 $x-z=0$．显然 $f$ 本身是非退化的二次型，但在由直线方程 $x=z$ 定义的 $\boldsymbol{R}^4$ 的 3 维子空间上成为退化的二次型且既取正值又取负值．我们用定理 1 来检验这一事实．

$$L=y^2+z^2-t^2-x^2+\lambda(x-z)$$

的驻点形式为 $v=(2t,t,0,t,0)$，$t\in\boldsymbol{R}$．计算 $L$ 的 $5\times5$ 阶 Hesse 矩阵得 $\Gamma_3=2$，$\Gamma_4=0$ 及 $\Gamma_5=0$．序列 2,0,0 是半正定的．从而定理 1 的 (b) 不能用于这种情形．但是交换 $z$ 和 $t$ 后得到 Lagrange 函数

$$y^2+t^2-z^2-x^2+\lambda(x-t)．$$

这时算得 $\Gamma_3=2$，$\Gamma_4=-4$ 及 $\Gamma_5=0$。因为序列 2，-4,0 是鞍型序列，所以由定理 1 的(b)′，可以断言，对于限制条件 $x=z$，$f$ 在形式为 $(t,0,t,0)\in \boldsymbol{R}^4, t\in \boldsymbol{R}$，的任意一个点上都不取极值。

### 3. 定理 1 的证明

我们首先对在 § 1 中引进的简单约束条件 $x_a=0$，$1\leqslant a\leqslant m$ 证明定理。针对这些约束条件分析 $HL(v)$ 将阐明带符号的子式 $\Gamma_{2m+p}$，$1\leqslant p\leqslant n-m$的作用。设

$$\tilde{f}(x_{m+1},\cdots,x_n)=f(0,\cdots,0,x_{m+1},\cdots,x_n)$$

相应的 Lagrange 函数 $L=f+\sum_{a=1}^{m}\lambda_a x_a$，其驻点是形式为

$$v=(c,a)\in \boldsymbol{R}^m\times U$$

的点，其中

(i) $a=(0,\cdots,0,a_{m+1},\cdots,a_n)$ 而 $\tilde{a}=(a_{m+1},\cdots,a_n)$ 是 $\tilde{f}$ 的驻点。

(ii) $c=\left(-\frac{\partial f}{\partial x_1}(a),\cdots,-\frac{\partial f}{\partial x_m}(a)\right)\in \boldsymbol{R}^m$.

显然有

$$\frac{\partial^2 L}{\partial x_i\partial x_j}(v)=\frac{\partial^2 \tilde{f}}{\partial x_i\partial x_j}(\tilde{a}),\quad m+1\leqslant i,j\leqslant n. \tag{3.1}$$

从而，将 $HL(v)$ 写成分块矩阵形式，有

$$HL(v)=\begin{bmatrix} O_{m,m} & I_m & O_{m,n-m}\\ I_m & B & C\\ O_{n-m,m} & C^T & H\tilde{f}(\tilde{a})\end{bmatrix},$$

其中 $H\tilde{f}(\tilde{a})=\left[\frac{\partial^2\tilde{f}}{\partial x_i\partial x_j}(\tilde{a})\right]$是 $\tilde{f}$ 在驻点 $\tilde{a}=(a_{m+1},\cdots,a_n)$ 处的 Hesse 矩阵，$O_{r,s}$是 $r\times s$ 阶零矩阵，$Ir$ 是 $r\times r$ 阶单位阵。设 $(\Delta_p)_{1\leqslant p\leqslant n-m}$ 是 Hesse 矩阵 $H\tilde{f}(\tilde{a})$ 的顺序主子式序列，$(S_k)_{1\leqslant k\leqslant n+m}$ 是 Hesse 矩阵 $HL(v)$ 的顺序主子式序列。由定义得

$$\Gamma_k=(-1)^m S_k,\quad 1\leqslant k\leqslant n+m.$$

由行列式的初等性质，容易证得

(i) $S_{2m}=(-1)^m$,

(ii) $S_{2m+p}=(-1)^m\Delta_p, 1\leqslant p\leqslant n-m.$ (3.2)

从而有

$$\Gamma_{2m+p}=(-1)^{2m}\Delta_p=\Delta_p,\quad 1\leqslant p\leqslant n-m,\qquad (3.3)$$

即序列 $(\Gamma_{2m+p})_{1\leqslant p\leqslant n-m}$ 和序列 $(\Delta_p)_{1\leqslant p\leqslant n-m}$ 重合。

关系式(3.3)解释了由 Mann [10] 引进的带符号的行列式序列。为了证明定理，我们看到

相对于约束条件 $x_\alpha=0, 1\leqslant\alpha\leqslant m$，$f$ 在$a\in A$ 处取（严格的）极小值或（严格的）极大值或不取极值的充分必要条件是 $\tilde{f}$ 在 $\tilde{a}=(a_{m+1},\cdots,a_n)$ 处分别取到（严格的）极小值或（严格的）极大值或不取极值。

(3.4)

设

$$\tilde{q}(h)=\frac{1}{2}\sum_{i,j=m+1}^{n}\frac{\partial^2\tilde{f}}{\partial x_i\partial x_j}(\tilde{a})h_ih_j$$

是与函数 $\tilde{f}$ 在驻点 $\tilde{a}$ 的值相连结的二次型(参阅§1的(1.2))应用命题1.1于 $\tilde{f}$，再应用经典的结果命题2.1(a)及引理2.1于二次型 $\tilde{q}$，我们发现，由前面的(3.3)式及(3.4)式，定理1的结论(a)和(b)证得。

剩下的是证明结论(b)′.由前面表出的矩阵 $HL(v)$ 的形式，可以明显看到，对于 $HL(v)$ 的最后 $(n-m)$ 行及列的任意一个置换 $\pi$，$HL(v)$ 左上角的 $2m\times 2m$ 阶主子式是不变的，但 $H\tilde{f}(\tilde{a})$ 变成 $H\tilde{f}(\tilde{a})(\pi)$。从而

(3.3式的证明说明序列 $(\Gamma_{2m+p}(\pi))_{1\leqslant p\leqslant n-m}$ 和序列 $(\Delta_p(\pi))_{1\leqslant p\leqslant n-m}$ 重合。 (3.5)

应用命题 1.1 于 $\tilde{f}$，又用命题 2.1(b)于二次型 $\tilde{q}$，可以发现，由前面的 (3.4) 式和 (3.5) 式,定理 1 的(b)′证得。

为了对于一般的约束条件证明定理 1，我们用坐标变换将其化为简单约束条件的情形。令

$$\varphi:\ \boldsymbol{R}^m\times U\to\boldsymbol{R}^m\times\boldsymbol{R}^n=\boldsymbol{R}^{m+n}$$

是映射

$$\varphi(\lambda_1,\cdots,\lambda_m,x_1,\cdots,x_n)=(\lambda_1,\cdots,\lambda_m,z_1,\cdots,z_n),$$

其中

$$z_a=g_a(x_1,\cdots,x_n),\quad 1\leqslant a\leqslant m,$$
$$z_{m+i}=x_{m+i},\quad m+1\leqslant i\leqslant n,$$

用分块矩阵的记号，有

$$\mathrm{D}\varphi(v)=\begin{bmatrix} I_m & O_{m,n}\\ O_{m,m} & \mathrm{D}g(a)\\ O_{n-m,2m} & I_{n-m}4\end{bmatrix}. \qquad (3.6)$$

由于 $\dfrac{\partial(g_1,\cdots,g_m)}{\partial(x_1,\cdots,x_m)}(a)\neq 0$，所以显然存在 $C^3$ 类的映射 $\varphi$，在 $v=(c,a)\in\boldsymbol{R}^m\times U$ 处有最大秩($=m+n$),因而由反函数定理[11]得,$(\lambda_1,\cdots,\lambda_m,z_1,\cdots,z_n)$ 在 $\boldsymbol{R}^{m+n}$ 中的 $w=\varphi(v)$ 的一个邻域内定义了一个坐标系。记

$$\varphi^{-1}(\lambda,z)=(\lambda,\psi(z))\in R^m\times U,$$

并设

$$\tilde{L}=L\circ\varphi^{-1}=f(\psi(z))+\sum_{a=1}^{m}\lambda_a z_a.$$

于是，$\tilde{L}$ 是对于约束条件 $z_a=0$，$1\leqslant a\leqslant m$，相应于 $C^3$ 类映射 $f\circ\psi$ 的 Lagrange 函数。由 $v$ 是 $L$ 的驻点得 $w=\varphi(v)$ 是 Lagrange 函数 $\tilde{L}=L\circ\varphi^{-1}$ 的驻点。显然，我们得到下面的等价条件：

对于约束条件 $g_a=0$，$1\leqslant a\leqslant m$，$f$ 在 $a\in A$ 点取到（严格的）极小值或（严格的）极大值或不取极值，当且仅当，对于简单约束条件 $z_a=0$，$1\leqslant a\leqslant m$，$f\circ\psi$ 在点 $a_0$ 处分别取到（严格的）极小值或（严格的）极大值或不取极值，其中 $\psi(a_0)=a$。

(3.7)

等价条件 (3.7) 指示我们应该比较 Hesse 矩阵 $HL(v)$ 和 $H\tilde{L}(w)$ 的顺序主子式序列。这由下面的引理来完成。

**引理3.1** $HL(v)=P^T\times H\tilde{L}(w)\times P$，其中 $P=D\varphi(v)$。

**推论3.1.1** 如果 $\pi$ 是 $\{1,\cdots,m+n\}$ 的一个置换，则 $HL(v)(\pi)=P^T(\pi)\times HL(w)(\pi)\times P(\pi)$。

**推论3.1.2** 另外，如果 $\pi$ 点态固定 $\{1,\cdots,2m\}$（即 $\pi$ 诱导 $(m+n)\times(m+n)$ 阶矩阵的最后 $(n-m)$ 行及列的一个置换），设 $(\tilde{\Gamma}_k(\pi))_{1\leqslant k\leqslant m+n}$ 是 $H\tilde{L}(w)(\pi)$ 的顺序主子式序列，则

$$\Gamma_{2m+p}(\pi)=\left[\frac{\partial(g_1,\cdots,g_m)}{\partial(x_1,\cdots,x_m)}(a)\right]^2\cdot\tilde{\Gamma}_{2m+p}(\pi),$$

$$0\leqslant p\leqslant n-m.$$

**推论3.1.3** 对于推论 3.1.2 中的每一个置换$\pi$，序列$(\Gamma_{2m+p}(\pi))_{1<p<n-m}$及序列（$(\widetilde{\Gamma}_{2m+p}(\pi))_{1<p<n-m}$分别都是正定（半正定）的、负定（半负定）的或鞍型的。

显然，推论 3.1.3 及前面的等式对于一般的约束条件证明了定理 1。

**引理 3.1 的证明** 这条引理是大家熟悉的结果，即坐标变换在相应的驻点处引出 Hesse 矩阵的一个相合矩阵的特殊情形。具体地说，因为$L=\widetilde{L}\circ\varphi$，所以在引进辅助记号

$$(u_1,\cdots,u_{m+n})=(\lambda_1,\cdots,\lambda_m,x_1,\cdots,x_n),$$

$$(v_1,\cdots,v_{m+n})=(\lambda_1,\cdots,\lambda_m,z_1,\cdots,z_n)$$

及

$$\varphi=(\varphi_1,\cdots,\varphi_{m+n}):\ \mathbf{R}^m\times U\to\mathbf{R}^{m+n}$$

后我们发现，由直接的链锁法则算得

$$\frac{\partial^2 L}{\partial u_i\partial u_j}(v)=\sum_{r,s=1}^{m+n}\frac{\partial^2\widetilde{L}}{\partial v_r\partial v_s}(w)\frac{\partial\varphi_r}{\partial u_i}(v)\frac{\partial\varphi_s}{\partial u_j}(v),$$

$$1\leqslant i,j\leqslant m+n,$$

这里，$v$是驻点，$w=\varphi(v)$。即

$$HL(v)=P^T\times H\widetilde{L}(w)\times P,\quad 其中\ P=\mathrm{D}\varphi(v).$$

**推论3.1.1的证明** 设$E$是相应于$\pi$的$(m+n)\times(m+n)$阶正交矩阵。于是有

$$\begin{aligned}HL(v)(\pi)&=E^{-1}\times HL(v)\times E\\&=E^{-1}\times P^T\times H\widetilde{L}(w)\times P\times E\\&=P^T(\pi)\times HL(w)(\pi)\times P(\pi).\end{aligned}$$

**推论3.1.2的证明** 由（3.6）式得，$P$的左上$2m\times 2m$阶主子矩阵是

$$P_{2m}=\begin{bmatrix} I_m & O_{m,m} \\ O_{m,n} & D_1(g)(a) \end{bmatrix},$$

其中

$$D_1(g)(a)=\left[\frac{\partial g_i}{\partial x_j}(a)\right]_{1\leq i,j\leq m}.$$

特别地，有

$$\det P_{2m}=\frac{\partial(g_1,\cdots,g_m)}{\partial(x_1,\cdots,x_m)}(a)\ (\neq 0).$$

如果$\pi$是$P$的最后（$n-m$）行及列的任意一个置换，则矩阵$P(\pi)$的形式是

$$P(\pi)=\begin{bmatrix} P_{2m} & B \\ O_{n-m,2m} & I_{n-m} \end{bmatrix}, \tag{3.8}$$

其中$B$是依赖$\pi$的$2m\times(n-m)$阶矩阵。

因为$HL(v)(\pi)=P^T(\pi)\times H\tilde{L}(w)(\pi)\times P(\pi)$，所以由(3.8)式（并注意到$P^T(\pi)=P(\pi)^T$）及简单的分块矩阵计算，推论3.1.2证得。从而，对于一般约束条件，定理1证得。

## 4. 讨论

这一节用于解释定理1与经典的结果，即§1的命题1.2之间的联系。首先我们将服从线性约束条件的二次型进行分类。

设$q=\frac{1}{2}\sum_{i,j=1}^{n}a_{ij}x_ix_j$是$\boldsymbol{R}^n$上的实值二次型，$A=[a_{ij}]_{1\leq i,j\leq n}$。假定$q$服从$m$个线性约束条件

$$\sum_{j=1}^{n}b_{ij}x_j=0,\quad 1\leq i\leq m;\ m<n.$$

令 $B=[b_{ij}]_{1\leq i\leq m;\ 1\leq j\leq n}$ 是这些约束条件的系数矩阵，并设 $m\times m$ 阶子矩阵 $[b_{ij}]_{1\leq i,j\leq m}$ 是非奇异的。显然，$v=(0,0)\in \boldsymbol{R}^m\times \boldsymbol{R}^n$ 是 Lagrange 函数

$$L=g+\sum_{i=1}^{m}\sum_{j=1}^{n}b_{ij}x_j\lambda_i$$

的驻点，在该点处的 Hesse 矩阵是

$$HL(v)=\begin{bmatrix}O_{m,m} & B\\ B^T & A\end{bmatrix}. \qquad (4.1)$$

由§2可知，$HL(v)$ 及 $HL(v)(\pi)$ 的顺序主子式序列分别是 $(\Gamma_k)_{1\leq k\leq m+n}$ 和 $(\Gamma_k(\pi))_{1\leq k\leq m+n}$ .

**定理2** 对于上面的线性约束条件，

(a) $q$ 是正(负)定的充分必要条件是序列

$$(\Gamma_{2m+p})_{1\leq p\leq n-m}$$

是正(负)定的。

(b) $q$ 是不定的充分必要条件是存在 $\boldsymbol{R}^n$ 中最后 $(n-m)$ 个变量的一个置换 $\pi$：$(x_1,\cdots,x_n)\to(x_1,\cdots,x_m,x_{\pi(m+1)},\cdots,x_{\pi(n)})$，使得序列 $(\Gamma_{2m+p}(\pi))_{1\leq p\leq n-m}$ 是非平凡的且是鞍型的。

(c) $q$ 是半正(负)定的充分必要条件是对于 $\boldsymbol{R}^n$ 中最后 $(n-m)$ 个变量的任意一个置换 $\pi$，序列 $(\Gamma_{2m+p}(\pi))_{1\leq p\leq n-m}$ 是平凡的或半正(负)定的。

**证明** 定理2的证明实质上是应用定理1于满足线性约束条件的二次型。首先假定这些线性约束条件是简单的：$x_i=0$，$1\leq i\leq m$。设 $\tilde{q}(x_{m+1},\cdots,x_n)$ 是二次型

$$q(0,\cdots,0,x_{m+1},\cdots,x_n)=\frac{1}{2}\sum_{i,j=m+1}^{n}a_{ij}x_ix_j.$$

在简单约束条件时，定理 2 由 § 3 的 (3.3) 式及 (3.5) 式（在定理 1 中令 $f=q$，$\tilde{f}=\tilde{q}$），由 § 2 的命题 2.1 给出的用于 $\tilde{q}$ 的分类结果及附录中的命题 4.1 推得。对于一般的线性约束条件，注意到如果 $f=q$ 及

$$g_a(x_1,\cdots,x_n)=\sum_{j=1}^{n}b_{aj}x_j,\qquad 1\leqslant a\leqslant m,$$

则在定理 1 中构造的坐标变换映射 $\varphi$：$\boldsymbol{R}^n\to\boldsymbol{R}^{m+n}$ 是线性映射。于是在新的坐标系中，函数 $q\circ\psi$ 是服从简单约束条件的二次型。从而，定理 2 由 § 3 的推论 3.1.3 及前面证明的相对于简单约束条件的二次型的分类结果证得。

**注释4.1** 结论 (a) 由 H. B. Mann 第一个建立[10]。结论(c)对于半定的二次型深刻且简化了 G.Debreu 获得的分类框架[5]（特别的，Debreu 的结果是用 $\boldsymbol{R}^n$ 中所有变量的置换叙述的）。定理 2 除了对 Lagrange 乘子法的理论明显有用外(参阅 § 1中的命题 1.1 和命题 1.2)，其中的结论(b)是新的。

现在对定理 1 中不能确定的情形进行研究，回想一下，§ 1 中的命题 1.2 是用 $\boldsymbol{R}^n$ 中的辅助二次型

$$q_L(h)=\frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2 L}{\partial x_i\partial x_j}(v)h_ih_j$$

叙述的，后者服从 $m$ 个线性约束条件(1.3)：

$$\sum_{i=1}^{n}\frac{\partial g_a}{\partial x_i}(a)h_i=0,\qquad 1\leqslant a\leqslant m.$$

**注意** 设 $L_1$ 是连结二次型 $q_L$ 及线性约束条件(1.3)的

Lagrange函数。由(4.1)式可以清楚地看到Hesse矩阵 $HL_1(0)$正好是§1中定义的Hesse矩阵 $HL_1(v)$。 (4.2)

由(4.2)及定理2，我们得到下面的结论：

(Ⅰ) 定理1的结论(a)成立，当且仅当对于约束条件(1.3)，二次型 $q_L$ 是正定的或负定的。

(Ⅱ) 定理1的结论(b)成立，当且仅当对于约束条件(1.3)，二次型 $q_L$ 是不定的。

(Ⅲ) 定理1中仅有的不确定的情形，在二阶导数的水平上当且仅当 $\det HL(v)=0$ 及(b)′不成立时出现，即，对于线性约束条件(1.3)，辅助二次型 $q_L$ 是退化的且是半正定或半负定的，有关这一点的数值算法由定理2的(c)提供。

(Ⅳ) 为完整起见，我们证明，上述不确定的情形，在二阶导数的水平上必定是不能确定的。

设 $V$ 是由线性约束条件(1.3)定义的 $\boldsymbol{R}^n$ 的 $(n-m)$ 维子空间。存在 $R^n$ 的一组基 $(e_1,\cdots,e_n)$ 使得 $(e_1,\cdots,e_{n-m})$ 是 $V$ 的一组基且对于这组基中的坐标 $(y_1,\cdots,y_n)$，二次型 $q_L$ 是对角型的：

$$q_L(y_1,\cdots,y_n)=\varepsilon(y_1^2+\cdots+y_k^2),$$

其中 $\varepsilon=1(\varepsilon=-1)$。如果在 $V$ 上，$q_L$ 是半正(负)定的且 $k$ 是 $q_L$ 在其上是正(负)定的 $V$ 的子空间的最大维数。由于 $q_L$ 在 $V$ 上是退化的，所以 $k<n-m$(如果 $k=0$ 则 $q_L=0$)。在上面提到的基中(假定 $m>0$)，线性约束条件(1.3)等价于简单约束条件：$y_i=0$，$i=n-m+1,\cdots,n$。对于简单约束条件(如果 $m>0$)：$y_i=0$，$i=n-m+1,\cdots,n$，令

$$f(y_1,\cdots,y_n)=q_L(y_1,\cdots,y_n)+y_{k+1}^4,$$

$$g(y_1,\cdots,y_n)=q_L(y_1,\cdots,y_n)-y_{k+1}^4.$$

对于这些简单约束条件，虽然二者的二次型都是$q_L$，但如果$\varepsilon=1$，则$f$在$\boldsymbol{R}^n$的原点处取最小值，但$g$不取极值；如果$\varepsilon=-1$，也有类似的结论。因此，不确定的情形(Ⅲ)必定是不能确定的。

**注释4.2** 由前面的(Ⅰ)和(Ⅱ)推得，定理1提供了一种数值算法，去区分§1中命题1.2的3种情形。

于是，我们完成了对定理1及条件极值问题的研究。

## 附录

在本附录中，我们复习线性代数的某些结果并证明§2的命题2.1(b)。设$Q$是$\boldsymbol{R}^r$上的实二次型，相伴矩阵是$M$(在某组基中)。对于$\{1,2,\cdots,r\}$的每一个置换$\pi$，矩阵$M(\pi)$的顺序主子式序列是$(y_j(\pi))_{1\leqslant j\leqslant r}$(对这个记号参阅§2)。设

$$(x,y)=\frac{1}{2}(Q(x+y)-Q(x)-Q(y))$$

是相伴的双线性型

**命题4.1** (a) $Q$是半正定的，当且仅当对于每一个置换$\pi$，序列$(y_j(\pi))$或者是平凡的，或者是半正定的。

(b) $Q$是半负定的，当且仅当对于每一个置换$\pi$，序列$(y_j(\pi))$或者是平凡的，或者是半负定的。

**证明** 只需证明(a)，因为对于(b)，只要考虑二次型$-Q$就行了.由半定二次型的传统分类(Debreu[5]，Gantmacher[7,p.307])得：$Q$是半正定的，当且仅当$M$的每一个$j$阶主子式$\geqslant0$，$j=1,\cdots,r$。对于置换$\pi$，注意到$M$的每一个

$j$阶主子式就是$M(\pi)$的左上$j$阶主子式，因此我们得到充分必要条件：$y_j(\pi)\geqslant 0$，$j=1,\cdots,r$，对任意一个置换$\pi$成立。另外，如果序列$(y_j(\pi))_{1\leqslant j\leqslant r}$是非平凡的，则只可能是+号序列后面跟着0序列。这是因为，由+号和0组成的任意其它非平凡序列都将是鞍型序列，因而由§2的引理2.1推得$Q$不是半正定的。由此看到，命题4.1深化了半定二次型的传统分类。

**引理4.1** 如果存在$\{1,2,\cdots,r\}$的2个置换：$\pi$和$\sigma$，使得序列$(y_j(\pi))_{1\leqslant j\leqslant r}$及$(y_j(\sigma))_{1\leqslant j\leqslant r}$是非平凡的且分别是半正定和半负定的，则存在$\{1,2,\cdots,r\}$的一个置换$\tau$，使序列$(y_j(\tau))_{1\leqslant j\leqslant r}$是非平凡的且是鞍型的。

**证明** 设$y_1(\pi)=Q(e)>0$，$y_1(\sigma)=Q(f)<0$，其中的$e$和$f$是$\boldsymbol{R}^r$的给定的一组基中的元素。设$\tau$是$\{1,2,\cdots,r\}$的一个置换，使矩阵$M(\tau)$的左上$2\times 2$阶主子矩阵是

$$\begin{bmatrix}(f,f) & (e,f)\\ (e,f) & (e,e)\end{bmatrix},$$

则显然有

$$y_1(\tau)=Q(f)<0,$$

$$y_2(\tau)=Q(f)Q(e)-[(e,f)]^2<0,$$

从而序列$(y_j(\tau))_{1\leqslant j\leqslant r}$是非平凡的且是鞍型的。

**命题2.1(b)的证明** 设$Q$是不定的，则$Q$既不是半正定的也不是半负定的。由命题4.1,或者存在一个置换$\pi$,使序列$(y_j(\pi))_{1\leqslant j\leqslant r}$是非平凡的且是鞍型的；或者存在置换$\pi$和$\sigma$满足引理4.1的假设。但无论哪一种情形都存在置换$r$使序列$(y_j(r))_{1\leqslant j\leqslant r}$是非平凡的且是鞍型的。因为§2的引理2.1证明了(c)中“当”的部分，所以命题2.1(b)证得。

# 参 考 文 献


[1] K.G. Binmore, Calculus, Cambridge University Press, 1983.

[2] F.Bowman *and* F.A. Gerard, Higher Calculus, Cambridge University Press, 1967.

[3] C.Carathéodory, Calculus of Variations and Partial Differential Equations, I, *Holden-Day*, 1965.

[4] R. Courant *and* F. John, Intrdcuction to Calculus *and* Analysis, Interscience Publishers, 1965.

[5] G. Debreu, Definite and semidefinite quadratic forms, Econometrica, 20 (1952), 295—300.

[6] C.H. Edwards, Jr., Advanced Calculus of Several Variables, Academic Press, 1973.

[7] F.R. Gantmacher, The Theory of Matrices, Vol.I, Chelsea, New York, 1977.

[8] H. Hancock, Theory of Maxima and Minima, Ginn, 1917.

[9] ,——Lectures on the Theory of Maxima and Minima of Functions of Several Variables, University of Cincinnati.

[10] H.B. Mann, Quadratic forms with linear constraints, Amer. Math. Monthly, 50 (1943), 430—433.

[11] J.Marsden *and* A.J. Tromba, Vector Calculus, 2nd ed., Freeman, 1976.

[12] Y. Murata, Mathematics for stability and Optimization of Economic Systems, Academic Press, 1977.

[13] A.L. Ostrosky *and* J.V. Koch, Introduction to Mathematical Economics, Houghton Mifflin, 1979.

[14] Ch. J. de la Vallée Poussin, Cours d'Analyse Infinitésimale, Dover Publications, New York, vol. 1, 8th. ed., Chapter I, Section 4 and Chapter IV, Section 3, 1946.


（朱学贤译，刘　勇校）

# 用穷竭法证明 $\ln 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$①

M. Finkelstein

$$\frac{1}{2}\left(\frac{2}{3} - \frac{2}{4}\right) = \frac{1}{3} - \frac{1}{4},$$

$$\frac{1}{4}\left(\frac{4}{5} - \frac{4}{6}\right) = \frac{1}{5} - \frac{1}{6}, \quad \frac{1}{4}\left(\frac{4}{7} - \frac{4}{8}\right) = \frac{1}{7} - \frac{1}{8},$$

$$\cdots\cdots$$

一般说来，有

$$\frac{1}{2^n}\left(\frac{2^n}{2^n + 2k - 1} - \frac{2^n}{2^n + 2k}\right) = \frac{1}{2^n + 2k - 1} - \frac{1}{2^n + 2k},$$

$$k = 1, 2, \cdots, 2^{n-1}; \quad n = 1, 2, \cdots.$$

因此（见图 1）有

$$\ln 2 = \int_1^2 \frac{\mathrm{d}x}{x} = \left(1 - \frac{1}{2}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \cdots$$

$$= 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots.$$

---

① $\ln 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$ (Proof by exhaustion), *Amer. Math. Monthly*, 94(1987), 541—542.

于是，我们可以下结论说，当 $x=1$ 时 $\ln(1+x)$ 等于它的 Maclaurin 级数。

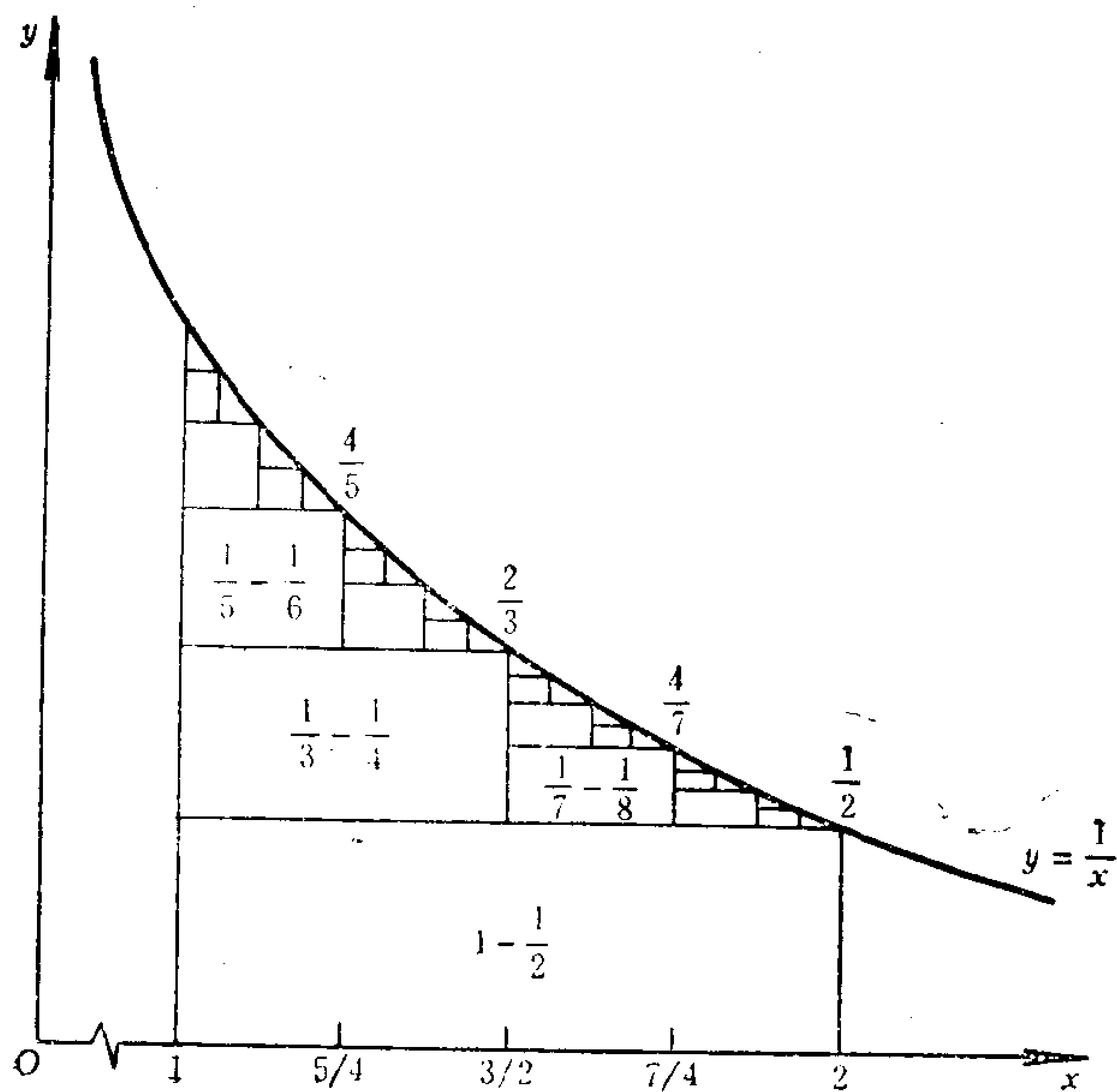



图 1

（朱学贤 译）

编者按　素数（也叫质数）是数学中最重要的基本概念之一。判断一个正整数是否是素数，及把一个正整数分解为素数的乘积问题，不仅是数论中的一个重要理论课题，而且具有重要的实用价值。这一问题的研究历史非常悠久，获得了丰富的成果，但仍远远没有解决。除了一些初等方法外，近年来研究这一看来十分简单明了的问题用到了很高深的近代数学方法，最近的进展可参看 H.W.Lenstra 的报告摘要：《数论中的有效算法》（数学译林，6(1987)，95—98)。

为了介绍这一重要课题，我们请张明尧同志翻译了 J.D.Dixon 在1984 年写的这篇综述性文章.文章介绍了研究这一课题的各种方法（主要介绍的是初等方法)、取得的结果，并用实例说明，很值得一读。有条件的读者可以利用计算机来实践这些方法。为了便于阅读，张明尧同志对相当部分的内容作了补充和改写，增加了文献[51]。与文章有关的初等数论、群、环、域、及概率论基础知识可参看通用的基础教科书。由于篇幅较长文章将分两次发表。

# 因子分解与素数判定① (一)

J.D.Dixon

## §1 引 言

初等数论中有一条关于整数可唯一分解的命题，它的严格表述如下：

设 $n>1$ 为 一个整数且有如下两个分解成素因子乘积的分解式

$$n=p_1\cdots p_r=q_1\cdots q_s,$$

那么必有 $r=s$，且在适当调换因子次序后，两表达式中各素因子依次相等.

这个定理也称为算术基本定理。这个定理很容易被看成是自然成立的，学生在学习初等数学时对它只有很肤浅的了解，因而很难了解它的深刻意义。历史上曾有许多重要的数学及计算方面的未解决的问题都与整数的分解及素数判定有关。正如L.E.Dickson 在其三卷本巨著《数论史》[12] ("History of the Theory of Number") 中所指出的，这些问题曾吸引了许多数学家的关注，其中有象 Fermat, Euler, Legendre 及 Gauss 这样一些著名数学家，计算机的行之有效对这方面的问题有良好积极的作用，它导致产生出一些算法，这些算

① Factorization and Primality Tests, *Amer. Math. Monthly*, 91(1984), 333—352.

法在以前是不可行的，它还对涉及基本问题的复杂性提出了有理论意义的新问题。

1978年，R. Rivest，A. Shamir 及 L. M. Adleman[50]提出了一种新型的公开密钥码，其可靠性在于某些大数很难分解这一假设。如有有效的因子分解方法，则将使敌方有破译此种密码之可能。

在这篇综述文章中，我的目的是对有关的问题及进展作一介绍。本文谈的主要是近十年来发展的成果。但读者应该明白，这一领域的历史要比这早得多，有些方法常被人以稍微不同的形式重新发现。有时很难确定一种思想有多少新东西，也很难弄清楚一种想法有多少成分包含在早先的（有时是久远得多的）研究成果之中。

## §2 环 $\boldsymbol{Z}_n$

设 $n\in \boldsymbol{Z}, n>1$，这里 $\boldsymbol{Z}$ 是全体整数组成的集合，$n$ 是我们对其因子分解感兴趣的一个整数。为了避免没有什么意义的特殊情形，我们总假设 $n$ 是奇数，且它有标准分解式

$$n=\prod_{i=1}^{s} r_i^{k_i},\quad 1\leqslant k_i\in \boldsymbol{Z}, \tag{1}$$

其中诸 $r_i$ 均为奇素数，且 $r_i\neq r_j\,(i\neq j)$。以下我们总用 $p, q$ 及 $r$（带下标或不带下标）来表示奇素数。

把全体整数按照被 $n$ 除后所得的余数来加以分类，余数相同的作为一类，称为以 $n$ 为模的一个剩余类，由于被 $n$ 除后所得余数（指非负余数）只有 $0,1,\cdots,n-1$ 这 $n$ 个，故以 $n$ 为模恰有 $n$ 个剩余类，它们分别记为 $\overline{0},\overline{1},\cdots,\overline{n-1}$。在每一类中取任一元作代表，这样得到的 $n$ 个整数称为一个完全

剩余系。在一个剩余类$\bar{a}$中，如果$a$与$n$互素，也就是$a$与$n$的最大公约数为1：

$$\mathrm{GCD}(a,n)=1,$$

则$\bar{a}$中任一整数必也与$n$互素（为什么?）。因此，我们就称$\bar{a}$是与$n$互素的剩余类。在所有与$n$互素的剩余类中各任取一数作为代表，这样得到的集合称为以$n$为模的一个简化剩余系，也称缩系。显然，同一剩余类中任何两个整数$a$与$b$被模$n$除有相同的余数，称为$a$与$b$关于模$n$同余，记为

$$a\equiv b \pmod n,$$

这也等价于说$n$整除$a$与$b$的差，即$n|(a-b)$。把每个剩余类看作一个元素，按照

$$\bar{a}+\bar{b}=\overline{a+b},$$

$$\bar{a}\cdot\bar{b}=\overline{ab}$$

来定义剩余类之间的加法及乘法。这样得到的剩余类集合记为$\boldsymbol{Z}_n$，它对上述加、乘法满足以下性质：

(1) 加法、乘法封闭：即对任意两个元$\bar{a},\bar{b}\in\boldsymbol{Z}_n$，也有$\bar{a}+\bar{b},\bar{a}\cdot\bar{b}\in\boldsymbol{Z}_n$。

(2) 加法满足结合律、交换律，

(3) 加法有单位元：事实上，对任一个$\bar{a}\in\boldsymbol{Z}_n$，皆有

$$\bar{0}+\bar{a}=\bar{a}+\bar{0}=\bar{a},$$

$\bar{0}$称为$\boldsymbol{Z}_n$的一个单位元。

(4) 加法有负元：对每个$\bar{a}\in\boldsymbol{Z}_n$，显然

$$\bar{a}+\overline{(-a)}=\overline{(-a)}+\bar{a}=\bar{0},$$

$\overline{(-a)}$ 称为 $\bar{a}$ 的负元，记为 $-\bar{a}$。

(5) 加法与乘法满足分配律：对任何 $\bar{a},\bar{b},\bar{c}\in Z_n$ 有

$$\bar{a}\cdot(\bar{b}+\bar{c})=\bar{a}\cdot\bar{b}+\bar{a}\cdot\bar{c}$$

$$(\bar{b}+\bar{c})\cdot\bar{a}=\bar{b}\cdot\bar{a}+\bar{c}\cdot\bar{a}.$$

(6) 乘法有单位元：对任何 $\bar{a}\in Z_n$，有

$$\bar{1}\cdot\bar{a}=\bar{a}\cdot\bar{1}=\bar{a},$$

$\bar{1}$ 称为 $Z_n$ 中一个单位元。

具有以上性质的集合称为一个环，所以 $Z_n$ 是一个环，它称为是以 $n$ 为模的剩余类环。此外它还满足关于乘法的交换律，这样的环称为交换环。

设 $\bar{a}\in Z_n$，如果存在 $\bar{b}\in Z_n$ 使

$$\bar{a}\cdot\bar{b}=\bar{b}\cdot\bar{a}=\bar{1},$$

那么就称 $\bar{b}$ 是 $\bar{a}$ 的一个（关于乘法的）逆元，记为 $(\bar{a})^{-1}$。让我们来看一个例子，

取 $n=9$。考虑其中 $\bar{2}$ 与 $\bar{3}$ 这两个元。容易看出

$$\bar{2}\cdot\bar{5}=\bar{5}\cdot\bar{2}=\overline{(2\times5)}=\overline{10}=\bar{1},$$

所以 $\bar{2}$ 以 $\bar{5}$ 为逆元，$\bar{5}$ 也以 $\bar{2}$ 为逆元。但由于

$$\bar{3}\cdot\bar{0}=\bar{0},\quad \bar{3}\cdot\bar{1}=\bar{3},\quad \bar{3}\cdot\bar{2}=\bar{6},$$

$$\bar{3}\cdot\bar{3}=\bar{0},\quad \bar{3}\cdot\bar{4}=\bar{3},\quad \bar{3}\cdot\bar{5}=\bar{6},$$

$$\bar{3}\cdot\bar{6}=\bar{0},\quad \bar{3}\cdot\bar{7}=\bar{3},\quad \bar{3}\cdot\bar{8}=\bar{6},$$

故 $\bar{3}$ 无逆元。

称 $\boldsymbol{Z}_n$ 中有逆元的元为单位数(unit)，$\boldsymbol{Z}_n$ 中所有单位数的全体记为 $U_n$，上例表明，一般来说 $U_n$ 不一定与 $\boldsymbol{Z}_n^*$ 相同（它表示 $\boldsymbol{Z}_n$ 中除去 $\bar{0}$ 元后的集合）。如果有某个 $n$ 使 $\boldsymbol{Z}_n^* = U_n$，即 $\boldsymbol{Z}_n^*$ 中每个元皆有逆元，则称这样的交换环 $\boldsymbol{Z}_n$ 是一个域。

我们现在要来讨论 $\boldsymbol{Z}_n$ 中元素 $\bar{a}$ 是单位数的条件。也即要问：对于什么样的 $\bar{a}$，才有整数 $x$，使

$$\bar{a}\cdot\bar{x} = \bar{1}? \tag{2}$$

上式等价于求同余方程

$$ax \equiv 1 \pmod n \tag{3}$$

的解。

如果 $x \equiv x_0 \pmod n$ 是(3)的解，由(3)式就有 $n \mid (ax_0 - 1)$，若 $\mathrm{GCD}(n,a) > 1$，则必有素数 $p, p\mid n, p\mid a$，于是设 $ax_0 - 1 = kn$，就也有

$$p \mid (ax_0 - kn) = 1,$$

但这不可能，故必有 $\mathrm{GCD}(n,a) = 1$。

反之，设 $\mathrm{GCD}(n,a) = 1$，我们要来证明：必有整数 $x_0$，$y_0$ 使

$$ax_0 + y_0 n = 1. \tag{4}$$

因为 $\bar{a} \in \boldsymbol{Z}_n$，我们总可以假设 $1 \leqslant a < n$。并考虑一切形如 $ax + yn$，$x, y \in \boldsymbol{Z}$ 的整数组成之集合 $T(a,n)$，其中必有一个最小的正数 $g = ax_0 + y_0 n \geqslant 1$。我们要证这个 $g$ 必为 $a$ 与 $n$ 之最大公约数，从而 $g = 1 = ax_0 + y_0 n$。

首先证 $a$ 与 $n$ 皆为 $g$ 的倍数。用反证法，若不然，必有两个整数 $u, v$ 使

$$a = gu + v, \quad u \geqslant 0, \ 1 \leqslant v < g. \tag{5}$$

于是

$$\begin{aligned} 1 \leqslant v &= a - gu = a - u(ax_0 + y_0 n) \\ &= a(1 - ux_0) - (uy_0)n \in T(a, n), \end{aligned}$$

但 $1 \leqslant v < g$，这与 $g$ 的最小性矛盾，故必 $v = 0$，即 $g|a$。同理可证 $g|n$。因而 $g$ 是 $a$ 与 $n$ 的一个（正的）公约数。

另一方面，由 $g = ax_0 + y_0 n$ 即知，$a$ 与 $n$ 的任一个公约数必整除 $g$，因而 $g = \mathrm{GCD}(a, n) = 1$。

由(4)式立即有：存在 $x_0 \in \boldsymbol{Z}$ 使 $n|(ax_0 - 1)$，也即存在 $\bar{x}_0 \in \boldsymbol{Z}_n$ 使 $\bar{a} \cdot \bar{x}_0 = \bar{1}$。我们就证明了如下的结论：$\bar{a}$ 为 $\boldsymbol{Z}_n$ 中单位数的充分必要条件是 $\mathrm{GCD}(a, n) = 1$。

对于剩余类的乘法，$U_n$ 满足乘法封闭性、结合律，乘法有单位元 $\bar{1}$，每个元有逆元，这样的集合 $U_n$ 称为一个群，由于它对乘法还满足交换律，故又称为交换群或 Abel 群。用 $|U_n|$ 表示群 $U_n$ 的阶，即它的元素个数，可以证明（读者自证）

$$|U_n| = \varphi(n). \tag{6}$$

这里 $\varphi(n)$ 是 Euler $\varphi$- 函数，它定义为不超过 $n$ 的正整数中与 $n$ 互素的正整数的个数，且我们有计算公式

$$\varphi(n) = n \prod_{i=1}^{s} \left(1 - \frac{1}{r_i}\right). \tag{7}$$

仍取 $n = 9$ 为例。由于 1,2,3,4,5,6,7,8,9 中与 $n$ 互素的恰有六个：1,2,4,5,7,8，因而

$$U_9 = \{\bar{1}, \bar{2}, \bar{4}, \bar{5}, \bar{7}, \bar{8}\},$$

这恰与由(6)及(7)式算出的阶数

$$|U_9| = \varphi(9) = 9\left(1-\frac{1}{3}\right) = 6$$

吻合。

设$G$是一个群，它上面给定一个乘法运算，因而可定义元素的乘幂。设$R$是$G$的一个（有限或无限的）子集，如果对任一元$g\in G$，皆可从$R$中找到某有限个元$a,b,\cdots,c$使

$$g = a^u b^v \cdots c^w, \quad u,v,\cdots,w\in \boldsymbol{Z},$$

则称$R$为$G$的一个生成集。若$R$是有限集，则称$G$是有有限生成元的群。特别当$|R|=1$，即$G$恰由一个元$a$生成时，称$G$为循环群，称$a$为$G$的生成元或本原元。当$G=U_n$时，若$U_n$为循环群，则称生成元为一个原根(mod $n$)。

关于何种$U_n$才有原根，初等数论中有如下重要的结论：设$m>1$为整数，则当且仅当$m$有下列形状之一时$U_m$才有原根存在，且原根恰有$\varphi(\varphi(m))$个：

$$m = 2, 4, p^\alpha, 2p^\alpha,$$

其中$p$为奇素数，$1\leqslant\alpha\in \boldsymbol{Z}$（参见文献[61]第六章）①。

特别地，对奇数$n>1$，仅当$n=p^\alpha$为素数幂时$U_n$才有原根，即此时$U_n$为循环群。

**习题 1** 证明$U_n$不能由少于$s$个元素生成($s$同式(1))。

下面要给出一条群论中的初等性质，它在本文后面要反复用到，为此先给出一些定义。

设$G$为一个群，$x\in G$，$G$的单位元记为$e$。若有正整数$h$使$x^h=e$，则称$x$为一个有限阶元素。使$x^h=e$成立的最

① 参看华罗庚：数论导引，第三章§§7,8,9.——译注

小正整数 $h$ 称为 $x$ 的阶。

**引理1** 设 $x$ 是群 $G$ 中一个 $h$ 阶元，$m$ 是一个正整数。又设 $x^m=e$ 且对某个素数 $q|m$ 有 $x^{m/q}\neq e$。又设 $t$ 为正整数，它使

$$q^t|m,\ q^{t+1}\nmid m \tag{8}$$

（以后上式简记为 $q^t\parallel m$）。则

(1) $q^t|h$

(2) 若对所有 $q|m$ 的素数皆有 $x^{m/q}\neq e$，则有 $m=h$（读者自证。由 $h|m$，$h\nmid m/q$ 即可推出——译注）。

最后，我们来回忆有关二次剩余的某些基本结果①。设 $a\in \boldsymbol{Z}$，$p$ 是一个素数且 $p\nmid a$。如果存在一个 $c\in \boldsymbol{Z}$ 使 $c^2\equiv a(\bmod p)$，则称 $a$ 是 $p$ 的一个二次剩余，反之则称 $a$ 是 $p$ 的一个二次非剩余。定义Legendre符号 $\left(\frac{a}{p}\right)$ 如下：

$$\left(\frac{a}{p}\right)=\begin{cases}0, & 若 p|a,\\ +1, & 若 a 为 p 的二次剩余,\\ -1, & 若 a 为 p 的二次非剩余.\end{cases} \tag{9}$$

Euler证明的一个重要结论（称为Euler判别法）是说，对任何奇素数 $p$ 及任何整数 $a$ 有

$$\left(\frac{a}{p}\right)\equiv a^{(p-1)/2}\quad (\bmod\ p). \tag{10}$$

对于任何奇数 $n>1$ 及任何与 $n$ 互素的整数 $a$，可以定义更一般的Jacobi符号 $\left(\frac{a}{n}\right)$ 如下：

---

① 参看华罗庚：数论导引，第三章。——译注

$$\left(\frac{a}{n}\right)=\prod_{i=1}^{s}\left(\frac{a}{r_i}\right)^{k_i}, \tag{11}$$

其中诸$\left(\frac{a}{r_i}\right)$均为Legendre 符号，而 $n$ 有（1）式之 标准分解式。

显然，Jacobi 符号取值为 1 或 − 1 。它还有以下一些简单性质。

引理2　设$n,n'>1$均为奇数，

(i) 如果$a\equiv b \pmod n$ 且GCD$(a,n)=1$，则有

$$\left(\frac{a}{n}\right)=\left(\frac{b}{n}\right). \tag{12}$$

(ii) 如果GCD$(a,n)$ = GCD$(a,n')=1$，则有

$$\left(\frac{a}{n}\right)\left(\frac{a}{n'}\right)=\left(\frac{a}{nn'}\right). \tag{13}$$

(iii) 如果GCD$(a,n)$ = GCD$(b,n)=1$，则有

$$\left(\frac{a}{n}\right)\left(\frac{b}{n}\right)=\left(\frac{ab}{n}\right) \tag{14}$$

注意，与Legendre符号不同的是，$\left(\frac{a}{n}\right)=1$只是同余方程

$$x^2\equiv a \pmod n \tag{15}$$

可解的必要条件，而不是充分条件。例：取 $n=49$，$a=3$。直接验算易知

$$1^2\equiv1,2^2\equiv4,3^2\equiv2,4^2\equiv2,5^2\equiv4,6^2\equiv1 \pmod 7,$$

故 3 是素数 7 的二次非剩余，由定义有$\left(\frac{3}{7}\right)=-1$。于是再

由定义得

$$\left(\frac{3}{49}\right)=\left(\frac{3}{7}\right)^2=(-1)^2=1.$$

下面只要证

$$x^2\equiv3 \pmod{49} \tag{16}$$

无解就说明问题了。如果上式有一解$x_0$，则有

$$49\mid(x_0^2-3),$$

从而$7\mid(x_0^2-3)$，这表明$x_0$也是

$$x^2\equiv3 \pmod{7} \tag{17}$$

的解，但上面已证明（17）不可能有解。

关于符号$\left(\frac{a}{n}\right)$，有如下基本的结果，它是初等数论中最重要的定理之一。

引理3　设$\mathrm{GCD}(a,n)=1$，$a>1$是奇数。

(i) $\left(\frac{-1}{n}\right)=(-1)^{(n-1)/2}$，

(ii) $\left(\frac{2}{n}\right)=(-1)^{(n^2-1)/8}$，

(iii)（二次互反律）

$$\left(\frac{a}{n}\right)=(-1)^{\frac{a-1}{2}\cdot\frac{n-1}{2}}\left(\frac{n}{a}\right).$$

有关证明请见[61]第V章。

## §3　有关计算的基本运算

在这一节里，我们要来分析一下执行各种算法所需的时

间。大多数计算机的硬件都能对某种固定位数的整数进行基本算术运算（单精度），其速率可达每秒 $10^{10}$ 次，但是处理更大整数的运算时则要用到（大概是由程序员所写的）多倍精度包（multiprecision package）。有关多倍精度包的详尽讨论见[21] §4.3.1。这里我们只需注意以下事实即可：对 $n$ 这么大小的整数作多精度运算要比大小为 $c(\ln n)^2 n$（这里 $c$ 为某个常数）的整数作单精度运算还要慢。通常我们把这一事实表述为“多精度运算是单精度运算的 $O((\ln n)^2)$ 倍”。其中记号 $O$（读作大 $O$，$O$ 按英文字母发音）定义如下：若当 $n\to\infty$ 时存在一个与 $n$ 无关的正常数 $c_1$ 使

$$|g(n)|\leqslant c_1 f(n)$$

对所有 $n$ 的取值恒成立，则记作 $g(n)=O(f(n))$ 或 $g(n)\ll f(n)$，其中 $g(n)$ 为实值函数，而 $f(n)$ 为一个取正实值的函数。

基本算术运算包括加法、减法、乘法及带余除法。于是 $\boldsymbol{Z}_n$ 中的环运算就有可供比较的时间。从微型机到最快的计算机计算一次单精度运算只需 $10^{-3}$ 秒到 $10^{-10}$ 秒左右的时间，这似乎是微不足道的。

除了四则运算之外，还有几种基本的数论算法（参见[21]及[28]），下面简要作一介绍。

1　广义欧几里得（Euclid）算法

设 $a$ 与 $b$ 为不全为 0 的两个整数，我们要计算它们的最大公约数 $d=\mathrm{GCD}(a,b)$，并求整数 $u,v$ 使 $au+bv=d$。设 $\mathrm{GCD}(a,n)=1$，由前述可求得 $u,v$ 使 $au+nv=1$，则 $au\equiv 1 \pmod n$，从而 $\bar{u}$ 就是 $\bar{a}$ 在环 $\boldsymbol{Z}_n$ 中的逆元。

设 $a$ 与 $b$ 是两个大小不超过 $n$ 的正整数，不妨设 $a\geqslant b$。

记$a=r_0$，$b=r_1$，则作辗转相除法就有

$$\begin{aligned} &r_0=r_1q_1+r_2, \quad 0\leqslant r_2<r_1, \\ &r_1=r_2q_2+r_3, \quad 0\leqslant r_3<r_2, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &r_{k-2}=r_{k-1}q_{k-1}+r_k, \quad 0\leqslant r_k<r_{k-1} \\ &r_{k-1}=r_kq_k. \end{aligned} \tag{18}$$

于是

$$r_k=\mathrm{GCD}(r_{k-1},r_k)=\mathrm{GCD}(r_{k-2},r_{k-1})=\cdots$$

$$=\mathrm{GCD}(r_0,r_1)=\mathrm{GCD}(a,b).$$

易见必有$r_k\geqslant 1$，$q_k\geqslant 2$．因若$q_k=1$，就有$r_{k-1}=r_k$，这与（18）中$r_k<r_{k-1}$矛盾．可以证明①（例如可参看[21] §4.5.2)：在上述辗转相除法中所需行带余除法的步数$k<C_2\ln a$，其中$C_2$为某个与$a$无关的正常数．

与上面类似地，通过反复运用Jacobi符号的性质（12）及互反律，从而化为引理3中的（i)(ii）两种情形，算出所给符号$\left(\frac{a}{n}\right)$的值．由于每用一次性质（12）实际上就是做一次带余除法．而用一次互反律后再用性质（12）恰相当于用上次带余除法的除数作为被除数，用上次带余除法的余数（若这余数是偶数，需分解出因子2的幂，将奇因子保留下来）作为除数．因而与上类似可证，算出$\left(\frac{a}{n}\right)$的值所需应用互反律及性质（12）的步数也不超过$\ln n$（设$n\geqslant a$）的一个常数倍．我们来看一个具体的例子．

① 见闵嗣鹤，严士健：初等数论，第一章§2，习题4.——译注

**例** 试计算$\left(\frac{383}{443}\right)$的值.

$$\left(\frac{383}{443}\right)\xlongequal{\text{互反律}}(-1)^{\frac{382}{2}\cdot\frac{442}{2}}\left(\frac{443}{383}\right)\xlongequal{(12)}-\left(\frac{60}{383}\right)$$

$$\xlongequal{\quad}-\left(\frac{2}{383}\right)^2\left(\frac{15}{383}\right)=-\left(\frac{15}{383}\right)$$

$$\xlongequal{\text{互反律}}(-1)(-1)^{\frac{14}{2}\cdot\frac{382}{2}}\left(\frac{383}{15}\right)=\left(\frac{383}{15}\right)$$

$$\xlongequal{(12)}\left(\frac{8}{15}\right)\xlongequal{\quad}\left(\frac{2}{15}\right)^2\left(\frac{2}{15}\right)=\left(\frac{2}{15}\right)$$

$$\xlongequal{\quad}(-1)^{\frac{15^2-1}{8}}=1.$$

2. 幂算法

对这一算法作更具一般性的描述是有用的. 设$R$是一个环, 取定一个非零元$x\in R$及一个正整数$m$, 要求$x^m$. 例如, 取$R=\boldsymbol{Z}_n$, 在一些重要的应用问题 (如大数分解、素数判别等) 中, 经常遇到计算$\bar{x}^m$即$x^m\,(\mathrm{mod}\ n)$的问题, 直接计算对大的$m$常需大的计算量. 如果把$m$写成二进位数的形式

$$m=m_0+m_1 2^1+\cdots+m_t 2^t, \tag{19}$$

其中每个$m_j$取0或1. 为了计算$x^m$, 可以先计算$x_0=x$, $x_1=x_0^2,\cdots,x_t=x_{t-1}^2$, 于是

$$x^m=x_{i_1}\cdots x_{i_k},$$

其中$i_1,\cdots,i_k$使$m_{i_1}=\cdots=m_{i_k}=1$, 而 (19) 中其余$m_j$皆为0. 这样总共只需要做至多$2t$次乘法运算即可算出$x^m$. 注意

到$2^t \leqslant m$，就有$t \leqslant (\ln m)/(\ln 2)$。于是，计算$x^m$只要$O(\ln m)$次乘法运算即可，而且不必对$m_i$及$x_i$作中间存储就可完成计算。这一算法显然和有时称做为“俄国农民乘法”的奇妙方法有关，这种乘法是用连续加倍来计算的。根据文献[21]§4.6.3所述，早在公元前200年以前的印度手稿中就曾出现过。幂算法在代数数域、多项式环、矩阵环以及在（用矩阵）求由常系数线性递推关系所给出的数列中的高次项的计算问题中也都是有用的。

## §4 概率算法

设$A$是一种确定算法，它的意义是从一个输入集$I$中接受一个元素$u$作为输入，经过有限时间$t(u)$的运算，产生一个结果$A(u)$，这个结果属于一个由某些可能的结果组成的指定集合。概率算法是这一概念的推广。所谓一个概率算法$P$，是从一个输入集$I$中取一个元素$u$作为输入，同时根据一种指定的概率分布从一个集合$\Omega_u$中得到一个元素$\omega$。在经过有限时间$t(u,\omega)$的运算后，这个算法以概率为1产生一个结果$P(u,\omega)$。此外，运算时间$t(u,\omega)$在$\Omega_u$上的平均值$t(u)$是有限的。如果一种概率算法是有用的话，其结果$P(u,\omega)$应有某种与$\omega$无关的性质，且$\Omega_u$的元素应易于通过正确的分布产生出来。概率算法多年来一直非正式地被人们使用着，不过关于它的第一个正式叙述（它与我们给出的解释略有差别）大概是由M.O.Rabin[47]给出的。给出正式定义的好处是使我们对这种算法可以进行严格的分析。注意到对给定的$\lambda>1$及一个固定的输入$u$，$t(u,\omega)$超过$\lambda t(u)$的概率小于$1/\lambda$，因此，估计$t(u)$对于算法$P$大概会运行多长时

间就有一个清楚的了解。在某些情形，运算时间严格有界是一项必不可少的重要条件，此时不能采用概率算法。而对几乎所有其它实际应用，使用概率算法与使用有可供比较的运算时间的确定算法有同样好的效果，且概率算法常常更容易加以分析。

在下面我们考虑的一些情形中，$\Omega_u$将是取自一个固定的有限区间中所有整数数列组成的集合。虽然$\omega$是一个无限序列，但我们只计算头几项，其中每一项都是以相互独立且一致的方式从一个给定的有限集中随机地挑选出来的（这就定义了$\Omega_u$上的概率分布）。而在实际计算中，我们并不是真的用随机抽取，而是多半用伪随机数发生器（参见[21]第3章）。

例 设给定一个素数$p$，要求它的一个二次非剩余。现在还不知道是否有一种确定算法，其运算时间不超过$\ln p$的一个幂次（比较[9]，但是请参见后面的引理2）。然而，通过用二次互反律来计算勒让德（Legendre）符号$\left(\frac{a}{p}\right)$或是应用Euler判别法（10）计算$a^{(p-1)/2}(\bmod p)$，很容易判断一个给定的整数$a$是否是$p$的二次非剩余。

若$p\geqslant 3$，容易看出，以下$(p-1)/2$个数

$$1^2,\ 2^2,\ \cdots,\ \left(\frac{p-1}{2}\right)^2$$

恰为模$p$的全部二次剩余。于是在模$p$的一个完全剩余系里，比方说在区间$[1,p-1]$的整数集中，恰有$(p-1)/2$个二次剩余及$(p-1)/2$个二次非剩余。我们可以用如下的概率算法来求$p$的一个二次非剩余：

相互独立且一致地从区间[1,$p-1$]中随机地选取一列整数作为待试的$a$,对每个取出的$a$计算$\left(\frac{a}{p}\right)$或$a^{(p-1)/2}\ (\bmod p)$，一旦对某个$a$计算结果为$-1$，则得到$p$的一个二次非剩余，计算即告终止。当然，结果得到的那个二次非剩余与所用的随机数列是有关的，即：使用不同分布的随机数列，所得到的可能是不同的二次非剩余。易见，平均说来只要试算两个$a$的值就可找到$p$的一个二次非剩余，因而，为了求得$p$的一个二次非剩余，平均来说要做$O((\ln p)^3)$次单精度运算。

**注** 因为对许多特殊类型的素数来说，求它们的二次非剩余有更简易的方法，故实际上我们很少会用到上述的算法。例如，由引理3的（i）与（ii）分别有

$$\left(\frac{-1}{p}\right)=-1,\quad p\equiv 3\ (\bmod 4)$$

$$\left(\frac{2}{p}\right)=-1,\quad p\equiv \pm 3\ (\bmod 8),$$

于是当$p\equiv 3\ (\bmod 4)$时，$a=-1$即为$p$之二次非剩余，而当$p\equiv 3$或$-3\ (\bmod 8)$时，$a=2$即为$p$之二次非剩余。又由二次互反律易有：对$p\equiv 17\ (\bmod 24)$

$$\left(\frac{3}{p}\right)=(-1)^{\frac{3-1}{2}\frac{p-1}{2}}\left(\frac{p}{3}\right)$$

$$=\left(\frac{p}{3}\right)=\left(\frac{17}{3}\right)$$

$$=\left(\frac{2}{3}\right)=-1,$$

因而当 $p\equiv17 \pmod{24}$ 时，取 $a=3$ 即为 $p$ 的一个二次非剩余。由于 $p\equiv3 \pmod 4$ 包含了形如

$$p\equiv7,11,19,23 \pmod{24}$$

的素数，$p\equiv\pm3 \pmod 8$ 包含了形如

$$p\equiv19,5,13 \pmod{24}$$

的素数，再加上形如 $p\equiv17 \pmod{24}$ 的素数，我们就对除去 $p\equiv1 \pmod{24}$ 以外的所有素数直接给出了相应的一个 二 次非剩余。对于剩下的素数，当然可以将上述方法扩展下去以对更多类型的素数直接写出它们相应的一个二次非剩余。

在因子分解及素性判别问题中，待测整数 $n$ 的大小是用量 $\ln n$ 来度量的，这是因为 $n$ 的大小可用 $n$ 的十进位表示的位数 $l$ 来度量，显然对 $l\geqslant1$ 有

$$10^{l-1}\leqslant n<10^l,$$

于是

$$\left(1-\frac{1}{l}\right)\ln 10\leqslant\frac{\ln n}{l}<\ln 10,$$

即位数 $l$ 与 $\ln n$ 除了一个常数倍外，基本上有同样的大小。一个主要的理论问题是：求解因子分解问题或是素性判别问题是否存在多项式算法？详言之，我们的问题如下述：是否有这样的算法（它也或许是概率算法），使得对于每个奇整数 $n>1$：

(i) 当 $n$ 是复合数时，我们能用此法求出 $n$ 的一个真因子；

(ii) 当 $n$ 是素数时，我们能用此法给出 $n$ 是素数的一个

证明。

而且，在无论哪一种情形，所需运算时间皆不超过 $\ln n$ 的某个幂次。

后面我们会看到，这些问题都是尚未解决的问题。已有的证据显示，问题(ii)可能有多项式算法，而对问题(i)似乎不存在多项式算法.同时，对于直到某种大小的一般性的 $n$，寻求合适的算法并执行这种算法以求解这些问题，仍有许多实际问题。到本文写作时，50位左右的复合数可按一定程序加以分解，而对 200位左右的素数可以有确定的方法证明它是素数。例如，利用L.M.Adleman和R.S.Rumely 1983 年提出的一种近似多项式算法的简化形式[11]，对 200 位左右的整数，可在 10 分钟左右判断出它是否是素数。

**习题2** 设 $p$ 与 $q$ 为素数，又设 $q-1=p^t m$，其中 $t\geqslant 1$ 且 $p\nmid m$。试给出一个概率算法来求 $U_q$ 中一个 $p^t$ 阶元 $b$。由此给出一个有效算法，以对每个 $a\in U_q$，用此算法来决定 $x^p=a$ 是否有解 $x\in U_q$，而如果有解的话，试求出一个解来。[提示：$a$ 是 $U_q$ 中一个 $p$ 次幂，当且仅当 $a^m$ 的阶整除 $p^{t-1}$。证明：在有解的情形，解 $x$ 可以写成 $a^{mi}b^j$ 的形式（请与[28]p.133 及[57]比较之）.]

**习题3** 如果 $g$ 是模 $p$ 的一个原根，$p$ 是一个奇素数，且 $g^{p-1}\not\equiv 1(\mathrm{mod} p^2)$。证明：对所有 $k>1$，$g$ 也是模 $p^k$ 的原根。注意，$g^{p-1}$ 与 $(g+p)^{p-1}$ 中至少有一个对模 $p$ 不同余于 1，于是，要么是 $g$，要么 $g+p$ 是模 $p^k$ 的一个原根。

**习题4** 设 $p$ 是一个素数，而 $p-1$ 可以完全分解。给出一个有效的概率算法，以对所有 $k\geqslant 1$ 求出模 $p^k$ 的一个原根。

## §5 素性判定

设 $n>1$ 为整数，那么要么 $n$ 是一个素数，要么它有一个素因子 $r\leqslant\sqrt{n}$。以这一事实为基础可以给出一个经典的算法，它既可以判别 $n$ 是否素数，又可在 $n$ 为合数时求出它的一个真因子。然而素数定理表明，小于 $\sqrt{n}$ 的素数个数渐近地趋向于 $2\sqrt{n}/\ln n$。因而当 $n$ 变得适当大时，要用这个判别法来判断 $n$ 是否素数及求合数 $n$ 的真因子，就是根本不可能实行的了。利用手头已有的计算资料，读者很可能会乐于考虑下面的问题。这个问题是在出版于 1907 年的一本谜题书[15]中提出来的。

**习题5** 数 $n=111\cdots1$（一共有 19 个 1）是素数吗？

为了用试除法来判别习题 5 中的数是否素数，我们需用不超过

$$n_0=\sqrt{\underbrace{111\cdots1}_{19个1}}$$

的所有素数来试除它，但 $n_0>10^9$，我们知道，不超过 $10^9$ 的素数就有多于五千万个，因而用试除法来做，计算量太大了。何况一般的计算机里只存储有小于 $10^8$ 的素数，因而无法较快地解决这个问题。

约在 1877 年前后，E. Lucas 和 T. Pepin 指出，在一些特殊情形，可以直接证明一个数是素数（见[12]p.376）。这个判别法仅当 $n-1$ 可以完全分解时才有用。我们把它叙述成定理 1。

**定理1**(Lucas) 设 $n$ 为自然数，且 $n-1$ 可完全分解，即

$$n-1=q_1^{s_1}\cdots q_t^{s_t},$$

其中诸 $q_i$ 为互不相同的素数，$s_i\geqslant 1, t\geqslant 1$ 为整数。若有一个整数 $a$ 使 $a^{n-1}\equiv 1 \pmod n$，且对每个 $q_i$ 皆有 $a^{(n-1)/q_i}\not\equiv 1 \pmod n$，则 $n$ 必为素数。[提示：先用引理 1 证群 $U_n$ 是一个阶为 $n-1$ 的循环群，$a$ 就是它的一个生成元。再由 $U_n$ 的阶为 $\varphi(n)$ 推出 $\varphi(n)=n-1$，由此证出 $n$ 为素数。]

我们把 Pepin 的判别法留作为一个习题。

**习题6** (Pepin)　设 $n=F_k=2^{2^k}+1 (k\geqslant 2)$ 为一个Fermat 数。如果 $n$ 是素数，证明 5 是 $n$ 的一个二次非剩余。证明此时 $n$ 的二次非剩余个数和原根个数相等，且任一个二次剩余必非原根，于是在 $n$ 为素数时 5 恰为 $n$ 的一个原根。由此证明：$n$ 为素数，当且仅当 $5^{(n-1)/2}\equiv -1 \pmod n$。

让我们来举两个应用的例子。

**例**　判别 4999 是否素数。

**解**　注意 $n-1=4998=2\cdot 3\cdot 7^2\cdot 17$。容易验证取 $a=3$ 时有

$$3^{4998}\equiv 1 \pmod{4999}.$$

又分别有

$$3^{4998/2}\equiv -1 \pmod{4999},$$
$$3^{4998/3}\equiv 2661 \pmod{4999},$$
$$3^{4998/7}\equiv 227 \pmod{4999},$$
$$3^{4998/17}\equiv 2420 \pmod{4999},$$

于是由 Lucas 判别法知 4999 必为素数。

**例**　判别 $n=F_4=65537$ 是否素数。

**解**　我们有

$$5^{65536/2}\equiv 625^{8192}\equiv(-2597)^{4096}$$

$$\equiv(-5902)^{2048}\equiv(33457)^{1024}$$
$$\equiv\cdots\equiv(256)^2$$
$$\equiv-1\pmod{65537},$$

于是由习题 6 知 $F_4$ 是素数。

定理 1 中的 $a$ 需要对所有 $q_i$ 适用，这个条件太苛刻了。J.L.Selfridge 对此作了改进，这就是下面的

习题7(Selfridge) 证明，奇整数 $n>1$ 为素数的充分必要条件是：对每个 $q_i\mid(n-1)$，皆存在一个整数 $a_{q_i}$，使

$$a_{q_i}^{n-1}\equiv1,\quad a_{q_i}^{(n-1)/q_i}\not\equiv1\pmod{n}.$$

以上判别法都要求 $n-1$ 能够完全分解，这在许多情形都难以实现。但是我们常可以给出 $n-1$ 的部分因子分解，是否可能利用这种分解给出某种素性判别法呢？首先我们介绍下面属于 F.Proth 的一个简单结果。

定理2 (Proth,1878) 设 $n>1$ 为奇数，$n-1=mq$，其中 $q$ 是一个奇素数且 $2q+1>\sqrt{n}$。又设有整数 $a$ 满足

$$a^{n-1}\equiv1,\quad a^m\not\equiv1\pmod{n},\tag{20}$$

则 $n$ 必为素数。[提示：先由(20)证出 $q\mid\varphi(n)$。由此利用 Euler $\varphi$-函数的计算公式(7)推出必有 $n$ 的一个素因子 $p_i$ 使

$$p_i\equiv1\pmod{2q}.\tag{21}$$

再由 $n\equiv1\pmod{2q}$ 得 $np_i^{-1}\equiv1\pmod{2q}$，于是

$$n\equiv p_i\pmod{2qp_i}.\tag{22}$$

若 $n\neq p_i$，由上式给出

$$n\geqslant p_i(2q+1)\tag{23}$$

由(21)及(23)式推出

$$n\geqslant(2q+1)(2q+1)>n,$$

矛盾。]

1914 年，H.C.Pocklington 得到了如下进一步的结果。

**习题** 8(Pocklington)　设给定整数 $n>1, n-1=ml$，其中因子 $m$ 已可完全分解。设对 $m$ 的每个素因子 $q$，皆存在一个整数 $a_q$，满足

$$a_q^{n-1}\equiv 1(\bmod n) \tag{24}$$

$$\mathrm{GCD}(a_q^{(n-1)/q}-1, n)=1. \tag{25}$$

证明，每个素数 $r|n$ 满足 $r\equiv 1(\bmod m)$。故当 $m\geqslant l$ 时，$n$ 为素数。

Pocklington 的结果是判别素数的一个有用的结果，但由于对 $n-1$ 的分解存在困难，故 $m\geqslant l$ 这个条件常不能满足。为此，1978年，D.H.Lehmer 与 Brillhart 等人给出了如下的判别法。

**定理**3(Lehmer-Brillhart)　设 $n>1$ 为奇数，它满足习题 8 中的条件。$k\geqslant 1$ 为整数，当 $k>1$ 时要求对 $\lambda=1,\cdots,k-1$ 有

$$(\lambda m+1)\nmid n. \tag{26}$$

又设 $u, v$ 是 $l$ 被 $2m$ 除所得的商及余数，即

$$l=2mu+v,\quad 1\leqslant v<2m. \tag{27}$$

而且有

$$n<(km+1)(2m^2+(v-k)m+1). \tag{28}$$

那么，$n$ 为素数的充分与必要条件是 $u=0$ 或者 $v^2-8u$ 不为完全平方数。[提示：证明 $n$ 为合数之充分必要条件是 $u\neq 0$ 且 $v^2-8u$ 为平方数。

(Ⅰ) 设 $n$ 是合数，由习题 8 可设

$$n=(cm+1)(dm+1),\quad c, d\geqslant k. \tag{29}$$

由此及（27）推出有

$$c+d\equiv v \pmod{2m}. \tag{30}$$

另一方面，由(28),(29) 及

$$cd\geqslant k(c+d)-k^2 \tag{31}$$

推出

$(km+1)(2m^2+(v-k)m+1)>(km+1)((c+d-k)m+1)$,

由此得

$$c+d-v<2m. \tag{32}$$

由（30）及（32）即可推出

$$c+d=v. \tag{33}$$

由 $l=\dfrac{n-1}{m}=cdm+c+d$ 及（27),(33) 可证

$$2u=cd\neq 0,$$

且有

$$v^2-8u=(c+d)^2-4cd=(c-d)^2.$$

（Ⅱ）现在设 $u\neq 0$，$v^2-8u=t^2$，则由 $n=ml+1$ 出发，利用（27）以及

$$2u=\frac{v^2-t^2}{4}$$

易得

$$n=\left(\frac{v-t}{2}m+1\right)\left(\frac{v+t}{2}m+1\right),$$

故 $n$ 为合数.]

有了定理 3，习题 5 即很容易解决。只要注意到对习题 5 的 $n$ 相应有

$$m = 3333330 = 2 \cdot 3^2 \cdot 5 \cdot 11 \cdot 37 \cdot 91,$$
$$l = 2mu + v, \quad u = 500000, \quad v = 3666667.$$

于是 $v^2 - 8u = 1344450488889$，易验证 $7 \mid (v^2 - 8u)$，但是 $7^2 \nmid (v^2 - 8u)$。相应取 $k = 1$ 即可证明习题 5 的 $n$ 是素数。

在某些情形，$n - 1 = ml$ 中完全分解出的部分 $m$ 还不够大，因而（28）不能满足。这时可利用 $m$ 的素因子下界帮助我们对 $n$ 是素数作出判断，下面就是这样的一个判别法。

**定理** 4(Lehmer-Brillhart) 设 $n$ 满足习题 8 的条件，又存在整数 $a$ 使

$$a^{n-1} \equiv 1 \pmod{n},$$

$$\mathrm{GCD}(a^{(n-1)/l} - 1, n) = 1.$$

假若 $m$ 的素因子至少大于 $M$，且

$$n < (Mm + 1)(2m^2 + (v - M)m + 1),$$

其中 $u$，$v$ 定义同定理 3 。那么，$n$ 为素数的充分必要条件是 $u = 0$ 或 $v^2 - 8u$ 不为完全平方数。

Lucas 判别法在域上的多项式环中有类似的结果。为了叙述方便，我们给出一些定义。

设 $p$ 为素数，则 $\boldsymbol{Z}_p$ 是以 $p$ 为模的剩余类环，它也是一个域，由 $\boldsymbol{Z}_p$ 中元素为系数作多项式，这种多项式的全体在通常多项式的加法及乘法运算下组成一个环，记为 $\boldsymbol{Z}_p[X]$。首项（即幂次最高的项）系数为 1 的多项式称为首一多项式。在 $\boldsymbol{Z}_p[X]$ 中可以考虑以一个固定多项式为模的同余式。则我们有如下结论。

**习题** 9（定理 1 的类似） 设 $f(X)$ 为 $\boldsymbol{Z}_p[X]$ 中一个首

一多项式。证明 $f(X)$ 可以表为 $Z_p[X]$ 中不同的首一不可约多项式之积。当且仅当

$$f(X)\mid(X^{p^d}-X)$$

时，$f(X)$ 的所有不可约因子多项式的次数都是 $d$ 的约数。由此证明，一个 $d$ 次首一多项式 $f(X)$ 是不可约的，当且仅当下二条件成立：

(1) $X^{p^d}\equiv X \pmod{f(X)}$,

(2) 对每个素数 $q\mid d$，皆有

$$X^{p^{d/q}}\not\equiv X \pmod{f(X)}$$

（后一判别条件可用环 $Z_p[X]/(f(X))$① 中的幂算法通过 $O(d\ln p)^3$ 次单精度运算加以检验）。因为在环 $Z_p[X]$ 中近似有 $\frac{1}{d}$ 的 $d$ 次首一多项式是不可约的，这个判别法可以作为一种概率算法用来构造 $Z_p$ 上指定次数的不可约多项式。也见[10]，[19]及[32]。）

如果 $n$ 是合数，只要将 $n$ 写成两个真因子的乘积并作乘法验证，就给出了一个简短的证明。E.T.Bell[5] 提到一件轶事，说的是1903年在美国数学会的一次会议上F.N.Cole就曾对 Mersenne 数 $M_{67}=2^{67}-1$ 确切地证明过它是合数。这样一个证明可以用包含 $O(\ln n)^2$ 次单精度运算的计算加以验证。当然，我们所要的是找到一个证明，这也是数学中共有的困难。

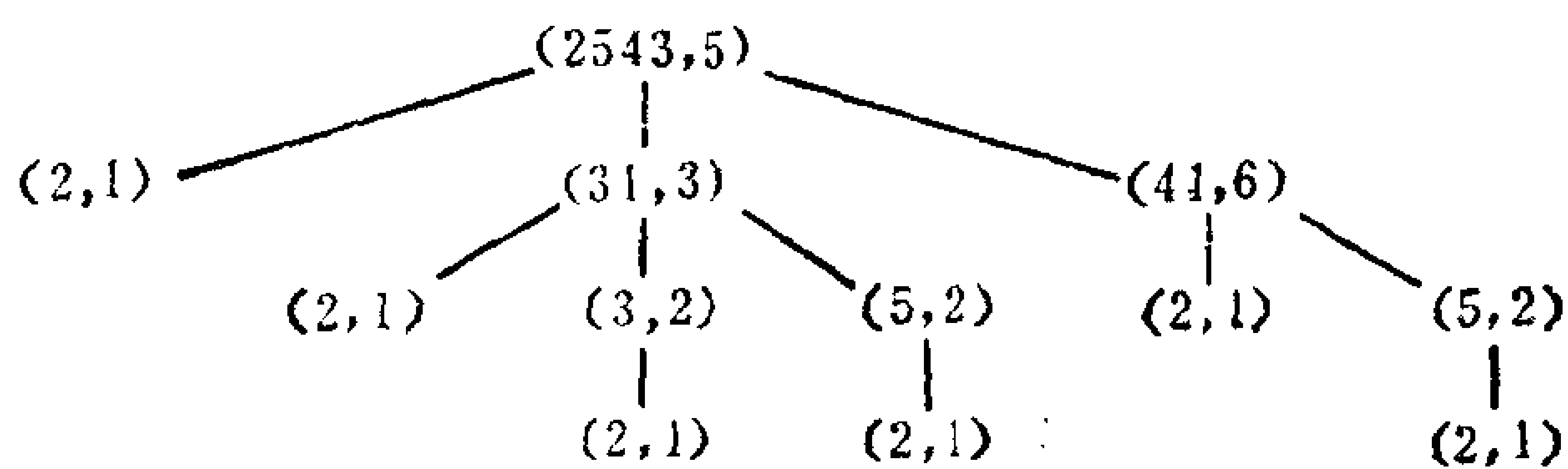
在 $n$ 是素数的情形，尽管在上一世纪就已对一些特殊情

---

① 这表示一般的剩余类环。参见张禾瑞：近世代数基础，第三章，§8。——译注

形给出了证明，但对检验素性是否有类似简短的证明存在，并不那么显然。然而，1975年V.P.Pratt[46] 注意到：Lucas判别法蕴含“每个素数均有简洁的判别法.”这种证明依旧可能很难发现，不过一旦将证明写了出来，其正确性很容易得到验证（Pratt 证明了：它可以用 $O(\ln n)^5$ 次单精度运算加以验证）。判别素数的证明可以写成一个有限树的形式，树的顶点标以数对 $(p,g_p)$，这里 $p$ 是一个素数，$g_p$ 是模 $p$ 的一个原根。树的根标以 $(n,g_n)$，每个顶点标以 $(p,g_p)(p>2)$，它以标签 $(q,g_q)$ 的顶点作为其子顶点，这里 $q$ 取遍整除 $p-1$ 的素数。树的叶均标以(2,1).

**例** $n=2543$ 的素性判别由下面的树给出，按照惯例，树是向下倒画的，树的叶在底下。




证明应该如下进行检验．对每个满足 $k>2$ 的顶点 $(k,g_k)$，我们应当有以下要求：(i) $g_k^{k-1}\equiv 1 \pmod k$，(ii) 只要当 $(l,g_l)$ 是顶点 $(k,g_k)$ 的一个子顶点的标签时，就有 $g_k^{(k-1)/l}\not\equiv 1 \pmod k$，(iii) $k$ 是当 $(l,g_l)$ 取遍顶点 $(k,g_k)$ 的诸子顶点时的诸 $l$ 的适当幂次的乘积．如果这些条件成立，则从叶(2,1)出发，Lucas 判别法就会依次证明：每个顶点的标签皆由一个素数及它的一个原根组成，特别推得 $n$ 是一个素数.

对于 Fermat 数 $F_6, F_7$ 与 $F_8$，R.P.Brent[7] 曾经对它们的素因子画出了象上面那样的证明树，但是因为需要对树中出现的所有素数给出 $p-1$ 的完全分解，因而很难给出这样的证明。另一方面，在[37]中指出了，可以较容易地人为造出一些很大的素数及其相应的证明树（从树的叶开始出发一直通到根部）。

问题仍然是：怎样找到一个素性证明？直到最近，这个问题也只好象比因子分解一个数的问题容易一点儿。习题 8 说明可以不要把 $n-1$ 完全分解。而 Lucas 曾指出，在某些情形可以用 $n+1$ 的分解来代替 $n-1$ 的分解。这个想法被 D.H.Lehmer 加以推广了（见[28]p.128），新近有人指出，可以用 $n^2\pm 1$ 及 $n^2\pm n+1$ 的部分分解合起来证明 $n$ 是素数。在 H.C.Williams[62] 中有关于这些方法的广泛深入的述评。目前看来，无论是从理论上还是从实用上讲，除了对象 Mersenne 数这样的特殊类型的整数外，由 Adleman，Rumely 及 C.Pomerance 于最近提出的算法应是最好的算法了，我们将在 §13 中讨论这一算法。

**注** 作为一项有趣的新结果我们注意到，在本文写作之时，已知最大的素数是 39751 位 Mersenne 数 $M_{132049}=2^{132049}-1$。这是由 D.Slowinski 于 1983 年 9 月证明的，其中用到 $n+1$ 是 2 的幂这一事实①。

## §6 伪素数与复合性证明

常常可以不用知道一个整数会有什么样的真因子就能证

① 最近证明了 $M_{216091}$ 也是素数。——译注

明它是一个复合数，这使许多人为之惊叹。如果 $n$ 是素数，那么 $Z_n$ 中所有单位数组成的群 $U_n$ 就是 $Z_n$ 中全体非零元的集合，即 $|U_n|=n-1$，于是，对任何整数 $b$，$n\nmid b$，就有

$$b^{n-1}\equiv 1 \pmod n, \tag{34}$$

这就是初等数论中著名的 Fermat 小定理。设对某个整数 $b$，有整数 $n>1$ 使(34)成立，我们就说 $n$ 通过了底为 $b$ 的伪素数检验，若此时 $n$ 为一复合数，则称 $n$ 为以 $b$ 为底的伪素数。因此，如果对某个 $b\in[1,n-1]$，$n$ 不能通过以 $b$ 为底的伪素数检验，我们就得到 $n$ 是合数的一个证明，这常常是证明复合性的一种简便有效的方法。

从历史上来说，关于 Fermat 小定理的逆命题有过一些含混错误的认识。如果(34)对某个 $b$ 及 $n$ 成立，是否 $n$ 一定为素数呢？据[12]p.91 所述，有一个时期 Leibniz 曾相信：凡通过以 2 为底的伪素数检验的整数皆为素数。容易看出，每个 Fermat 数 $F_k=2^{2^k}+1$ 都能通过以 2 为底的伪素数检验，可能正是这个事实引导 Fermat 错误地认为一切 $F_k$ 皆为素数。1732年，Euler 首先给出 $F_5$ 的分解

$$F_5=2^{2^5}+1=641\cdot 6700417,$$

从而推翻了 Fermat 的猜想。关于 $F_5$ 是合数的证明，也可按下题的方法做。

**习题10** 通过证 $F_5$ 不能通过底为了的伪素数检验来证明 $F_5$ 是合数。

到目前为止，人们已经发现许多 Fermat 数是复合数，而除了 $F_0,F_1,F_2,F_3,F_4$ 以外，迄今尚未找到新的 $F_k$ 是素数。

故现今有人猜想诸 $F_k$ 中可能仅有有限个素数存在。

1819年，法国的 F.Sarrus 指出

$$2^{341}\equiv 2 \pmod{341},$$

但是 $341=11\cdot 31$ 是合数。事实上，对每个底 $b>1$，都有无限多个复合数 $n$ 满足(34)。此即下述的定理5。

**定理5** 对每个整数 $b>1$，存在无穷多个以 $b$ 为底的伪素数。[提示：设 $p$ 是一个满足

$$p \nmid b(b^2-1) \tag{35}$$

的奇素数，作自然数

$$n=\frac{b^{2p}-1}{b^2-1}. \tag{36}$$

用(36)式直接证明 $n$ 是合数。由

$$(b^2-1)(n-1)=b\cdot(b^{p-1}-1)(b^p+b), \tag{37}$$

$$p(b^2-1)\mid(b^{p-1}-1), \tag{38}$$

$$2\mid(b^p+b) \tag{39}$$

推出

$$2p(b^2-1)\mid(b^2-1)(n-1),$$

即 $2p\mid(n-1)$。令 $n=1+2up$，利用

$$b^{2p}\equiv 1 \pmod{n} \tag{40}$$

推出 $b^n\equiv b \pmod{n}$，由 $p$ 的无限性即得欲证之结论。]

虽然对每个 $b>1$，以 $b$ 为底的伪素数有无穷个，但对每个固定的 $b>1$，在同一区间中以 $b$ 为底的伪素数与其中的素数相比，还是要少得多。例如，C.Pomerance[42] 中曾对不超过 $x$ 的伪素数个数 $\mathscr{P}(x)$ 给出以下的数值表

| $x$ | $\mathscr{P}(x)$ |
|---|---|
| $10^9$ | 5597 |

| | | |
|---|---|---|
| $5\cdot10^9$ | 11108 | |
| $10^{10}$ | 14884 | |
| $15\cdot10^9$ | 17658 | (41) |
| $2\cdot10^{10}$ | 19865 | |
| $25\cdot10^9$ | 21853, | |

他还证明了，当 $x$ 充分大时有

$$\mathscr{P}(x)<xL(x)^{-\frac{1}{2}}, \tag{42}$$

其中

$$L(x)=\exp\{\ln x\ \ln\ln\ln x/\ln\ln x\}. \tag{43}$$

（有关较早的结果见[16]）。而不超过 $10^9$ 及 $2.10^{10}$ 的素数分别有 50847534 及 882206716 个。又由 J.B.Rosser 与 L. Schoenfeld[51]中所给公式知，对 $x\geqslant17$ 有

$$\pi(x)>x/\ln x, \tag{44}$$

这里 $\pi(x)$ 表示不超过 $x$ 的素数个数。这些数值及公式均说明伪素数与素数相比个数要少得多。

1909年左右，R.D.Carmichael 证明了，存在合数 $n$，对任何 $b>1$，$(b,n)=1$，$n$ 都是以 $b$ 为底的伪素数。这样的合数称为 Carmichael 数。$n=561$ 是第一个被卡米凯尔发现的 Carmichael 数，也是最小的 Carmichael 数. 1912年,Carmichael 证明了如下结果。

**习题11** 假设 $n$ 是合数，其标准分解式是(1)。证明,对所有与 $n$ 互素的 $b$ 有(34)式成立的充分必要条件是：对每个 $i$ 有 $k_i=1$ 及 $(r_i-1)\mid(n-1)$。(后一条件蕴含 $s\geqslant3$，即 $n$ 至少有三个不同的素因子。)试找出一些是 Carmichael 数的例子。D.Shanks 发现，如果 $p>3$，$2p-1$ 以及 $3p-2$ 皆为素数，那么它们的乘积是一个 Carmichael 数)[提示：(Ⅰ)

充分性：对任意$b>1$，$(b,n)=1$，由Fermat小定理及$GCD(r_1-1,\cdots,r_s-1)\mid(n-1)$证出

$$b^{n-1}\equiv 1\ (\mathrm{mod}\ r_i)\ (i=1,\cdots,s),$$

故推出$n$为Carmichael数。（Ⅱ）必要性：设$n$为Carmichael数，有(1)形之标准分解。

① 证明$2\nmid n$。

先设$n=2^t$，$t>1$，取$b=3$，若

$$b^{n-1}\equiv 1\,(\mathrm{mod}\ n) \tag{45}$$

则由

$$b^n\equiv 1\ (\mathrm{mod}\ n) \tag{46}$$

与(45)立即推出$3\equiv 1(\mathrm{mod}\ n)$，故$t=1$，矛盾，于是(45)不能成立，这与$n$为Carmichael数矛盾。再设$n$是偶数且有一个奇素因子$r$。取$b$为$r$的一个奇的原根，由$b^{n-1}\equiv 1\ (\mathrm{mod}\ r)$得$(r-1)\mid(n-1)$，这与$n$为偶$r$为奇矛盾。

② 由①的证明知可设(1)式中诸$r_i$为奇素数。设$g_i$是模$r_i^{k_i}$的原根，由孙子定理可求得$b$适合

$$b\equiv g_i\,(\mathrm{mod}\ r_i^{k_i}),\quad i=1,\cdots,s。$$

故$GCD(b,n)=1$。由

$$b^{n-1}\equiv 1\,(\mathrm{mod}\ r_i^{k_i}),$$

$$b^{n-1}\equiv g_i^{n-1}\,(\mathrm{mod}\ r_i^{k_i})$$

及原根定义得$\varphi(r_i^{k_i})\mid(n-1)$，由此推出$k_i=1$，$i=1,\cdots,s$。又即推出$(r_i-1)\mid(n-1)$，$i=1,\cdots,s$。

③ 最后证$s\geqslant 3$。反证，若$s<3$，由$n$为合数有$s=2$。由$(r_1-1)\mid(r_1r_2-1)=n-1$推出$(r_1-1)\mid(r_2-1)$，同理$(r_2-1)\mid(r_1-1)$，故$r_1=r_2$，矛盾。

④ 2821 = 7·13·31, 10585 = 5·29·73,  27845 = 5·17·29·23, 172081 = 7·13·31·61等皆为 Carmichael 数。]

Carmichael 数的存在说明，在某些情形很难利用形如(34)的伪素数检验来找到素性①的证明，虽然尚不知道Carmichael 数究竟有多少，但一般人们猜测它们有无穷多个（例如见[42]）。为了判别素数，用稍微强一些的检验法会使效果好得多。记 $n-1=2^h m$，$2\nmid m$。如果 $n$ 是素数，则 $U_n$ 是循环群且有一个由 $-1$ 生成的二阶子群。同样，对每个 $x\in U_n$，$x$ 的阶整除 $n-1$，因而 $x^m$ 的阶整除 $2^h$。因此，要么 $x^m=1$，要么对某个 $i\in[0,h-1]$，$x^{m2^i}$ 的阶为 2，于是 $x^{m2^i}=-1$。所以，只要 $n$ 是素数且 $n\nmid b$，就有

$$\left.\begin{aligned} &b^m\equiv 1 \pmod n \\ \text{或}\quad & \\ &b^{m2^i}\equiv -1 \pmod n \quad (\text{对某个 } i\in[0,h-1]) \end{aligned}\right\} \qquad (47)$$

成立。显然(47)蕴含(34)。如果(47)成立，就说 $n$ 通过以 $b$ 为底的强伪素数检验。又如果 $n$ 为合数，则称 $n$ 是以 $b$ 为底的一个强伪素数。注意，虽然用幂算法的检验更严一些，但用幂算法检验(47)和直接检验(34)所需工作量大致相当。下面的定理 6 表明，对(47)这一检验法不存在象 Carmichael 数那样的类似物，在[44]中对强伪素数检验法的效力举了计算例证，这些例子表明，每个小于 $25\cdot10^9$ 且对 $b=2,3,5,7$ 满足(47)的奇数 $n$ 皆为素数。定理 6 的证明是以下面一条关于 $U_n$ 构造的结果为基础的。

① 原文此处误为复合性.——译注

**引理4** 设 $n>1$ 为奇数，$\langle u\rangle$表示 $U_n$ 中元素 $u$ ① 所生成的循环子群。设 $B$ 由 $U_n$ 中所有使$\langle u\rangle$的阶为奇数或$\langle u\rangle$中包含 $-1$的那种元素 $u$ 组成。如果 $B$ 生成 $U_n$，则 $n$ 是一个素数幂。

**证明** 设(1)为 $n$ 的标准分解。记 $n_i=r_i^{k_i}$ $(i=1,\cdots,s)$，对诸剩余类环 $Z_{n_i}$，作一个新集合 $\prod_{i=1}^{s} Z_{n_i}$，它的元形如 $(x_1,\cdots,x_s)$，$x_i\in Z_{n_i}$。定义其中两个元的加法与乘法分别为

$$(x_1,\cdots,x_s)+(y_1,\cdots,y_s)=(x_1+y_1,\cdots,x_s+y_s),$$
$$(x_1,\cdots,x_s)(y_1,\cdots,y_s)=(x_1y_1,\cdots,x_sy_s),$$

容易验证，对于这样定义的加法与乘法，$\prod_{i=1}^{s} Z_{n_i}$也作成一个环，它称为 $Z_{n_1},\cdots,Z_{n_s}$的直积。对每个元 $x\in Z_n$，定义一个对应关系 $f$，它把 $x$ 变为$\prod_{i=1}^{s} Z_{n_i}$中一个元$(x_1,\cdots,x_s)$，其中

$$x_i\equiv x \pmod{n_i}.$$

$f$ 显然满足：对任何 $x,y\in Z_n$，

$$f(x+y)=f(x)+f(y),$$
$$f(xy)=f(x)\cdot f(y).$$

这种 $f$ 称为环 $Z_n$ 到环$\prod_{i=1}^{s} Z_{n_i}$之间的一个同态，记为 $Z_n \overset{f}{\sim} \prod_{i=1}^{s} Z_{n_i}$。如果对每个$(x_1,\cdots,x_s)\in\prod_{i=1}^{s} Z_i$，必有 $x\in Z_n$ 使

① 以后为简单计，剩余类环中元素$\bar{x}$均记为$x$。——译注

$f(x)=(x_1,\cdots,x_s)$，则称 $f$ 是满同态。由于诸 $n_i$ 两两互素，由初等数论中著名的孙子定理（也称中国剩余定理）知，对任一个 $(x_1,\cdots,x_s)\in\prod_{i=1}^{s}\boldsymbol{Z}_{n_i}$，必有唯一的 $x\in\boldsymbol{Z}_n$ 使

$$x\equiv x_i \pmod{n_i},\quad i=1,\cdots,s。$$

因此上述同态不但是满的，还是一对一的，这样的同态称为一个同构。记为

$$\boldsymbol{Z}_n\overset{f}{\simeq}\prod_{i=1}^{s}\boldsymbol{Z}_{n_i}。\tag{48}$$

由此不难推出，也有

$$U_n\simeq\prod_{i=1}^{s}U_{n_i}。\tag{49}$$

上式表明，$U_n$ 中每个元素的阶整除诸 $\varphi(n_i)\ (i=1,\cdots,s)$ 之最小公倍数。记

$$\mathrm{LCM}\{\varphi(n_1),\cdots,\varphi(n_s)\}=2^t l,\quad 2\nmid l。\tag{50}$$

选取 $j$ 使 $2^t|\varphi(n_j)$。对每个元 $x\in U_n$，定义

$$\psi(x)=x^{2^{t-1}l},$$

显然 $\psi$ 是 $U_n$ 到 $U_n$ 的一个群同态，即对 $U_n$ 中乘法而言有

$$\psi(xy)=\psi(x)\psi(y),\ 对任何\ x,y\in U_n。$$

对每个 $x\in B$，易见 $\psi(x)\in\langle x\rangle$ 且

$$\psi(x)^2=(x^{2^{t-1}l})^2=x^{2^t l}=1。$$

由是，要么 $\psi(x)=1$，要么 $\psi(x)=-1$（$-1$是$\langle x\rangle$中唯一的二阶元）。

如果$B$生成$U_n$，则对所有$x\in U_n$有$\psi(x)=\pm 1$。另一方面，若$n$不是素数幂，那么(49)中就有$s>1$，因此$U_n$就包含一个元素$v$，它使$v \bmod n_j$在$U_{n_j}$中的阶为$2^t$，且对所有$i\neq j$有$v \bmod n_i=1$。显然$\psi(v)\neq\pm 1$。所以我们断言：若$B$生成$U_n$，那么$s=1$且$n$是素数幂，明所欲证。

**定理6** (a) 设$n-1=2^h m$，$2\nmid m$，定义集合

$B'=\{x\in U_n \mid x^m=1$ 或对某个 $i\in[0,h-1]$ 有 $x^{m2^i}=-1\}$。

如果$B'$生成$U_n$，则$n$为素数。

(b) (M.O.Rabin[47])。如果$n$为合数，那么(47)至少对$b$取$[1,n-1]$中的一半整数不成立。

**证明** (a) 用引理4的记号，我们有

$$B'\subseteq B$$

且对所有$x\in B'$有$x^{n-1}=1$。故若$B'$生成$U_n$，引理4表明对某个素数$r$有$n=r^k$，且对所有$x\in U_n$有$x^{n-1}=1$。而那就推出$U_n$是一个阶为$r^{k-1}(r-1)$的循环群，故$r^{k-1}(r-1)\mid(r^k-1)$，这就得出$k=1$，如所欲证。

(b) 由于$B'$由属于区间$[1,n-1]$且使(47)成立的那些元素$b \bmod n$所组成，(a)表明$B'$生成$U_n$的一个真子群。于是$|B'|\leqslant\frac{1}{2}|U_n|$，得证。

**注** 实际上Rabin用的不是(47)，而是它的一种更复杂、但是等价的形式。在[49]中他指出，在那种形式下，定理6中的"一半"可以换成"1/4"，他给出一些复合数$n$的例子，这些例子表明1/4已是本质上来说最好可能的了。

定理6(b)中所证明的性质可以用来给整数的复合性加以检验（在[47]及其它一些地方述及可用来作素性判别，这

是不对的)。固定一个整数 $k \geqslant 1$，相互独立且一致地从区间 $[1, n-1]$ 中随机选取 $k$ 个整数。若 $n$ 未能通过以这些数为底的强伪素数检验，则断定 $n$ 为合数，反之则得不到结论，但有相当证据显示 $n$ 极可能是素数。当第二可能出现，即 $n$ 通过以这些数为底的强伪素数检验时，对此始终有某种糊涂的看法。如 Rabin 所强调的，说"$n$ 可能是素数"（即 $n$ 或是素数，或不是素数），这毫无意义。对 $k$ 的无论什么值，这个概率算法也不能证明 $n$ 是素数。另一方面，若 $n$ 为合数，则此检验法判出 $n$ 是复合数的概率与 $n$ 无关，且至少为 $1-2^{-k}$（如果利用更强的结果，这个概率至少达 $1-4^{-k}$）。于是，若以 $k=100$ 为例，那么这个检验法极不可能判断不出 $n$ 是合数。因而，如果对于 100 组随机抽取的基，$n$ 都通过了强伪素数检验，那么 $n$ 极可能是一个素数，不过仍然不算是给出了证明。注意，每做一次 (47) 的检验需要做 $O(\ln n)^3$ 次单精度运算。

大约在与[47]发表的同时，R. Solovay 和 V. Strassen[60] 独立地提出了一种稍微不同的判别合数的概率方法(也见 D. H. Lehmer 的论文[29])。根据这个检验法，我们对随机选取的整数 $a \in [1, n-1]$ 来计算 Jacobi 符号 $\left(\frac{a}{n}\right)$ 的值，并将它与 $a^{(n-1)/2} \bmod n$ 加以比较。如果 $n$ 是素数，这两个值是同样的。然而，L. Monier [34a] 指出，Solovay-Strassen 的方法要比 Rabin 的方法差：Solovay-Strassen 的方法需要的计算量更大，且有时在探查非素数时效果较差。更近期的一些文章 [1]，[24]，[44]及其它一些都对有关的检验法作了叙述。

可否用某种方式来强化定理 6，以求得可用于证明素性

的结果呢？这个问题和在不知道 $n$ 的素因子时弄清何时 $U_n$ 的一个子集才生成 $U_n$ 差不多同样困难，这是一个非常困难的问题。如果假设一个未证明的著名猜想——广义 Riemann 猜想成立，则上面的问题可得到某种解决。

以下我们用缩写字 GRH (Generalized Riemann Hypothesis) 表示广义 Riemann 猜想。我们假设读者了解级数论中有名的调和级数

$$\sum_{n=1}^{\infty}\frac{1}{n}=1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\cdots.$$

这个级数是发散的，它的和是 $+\infty$。Euler 和 B. Riemann 等人先后研究过如下更一般的级数

$$\sum_{n=1}^{\infty}\frac{1}{n^s}, \tag{51}$$

这个级数对满足 $s=\sigma+it$，$\sigma>1$ 的复数 $s$ 都是收敛的，因而它对 $\mathrm{Re}s=\sigma>1$ 定义了一个以 $s$ 为复变数的函数，称为 Riemann $\zeta$-函数，记为 $\zeta(s)$。Riemann 于1859证明了 $\zeta(s)$ 可以解析开拓成全复平面上除 $s=1$ 外均解析的函数，除了一些显然零点外，$\zeta(s)$ 有无穷多个复零点 $\rho_n=\beta_n+i\gamma_n$ 在条形域 $0\leqslant\mathrm{Re}s\leqslant1$ 中。他猜测这些零点满足 $\beta_n=\frac{1}{2}$，这就是当今著称的 Riemann 猜想。

设 $\chi$ 是以 $m>1$ 为模的特征函数① （见[20a]第十六章），Dirichlet 首先考虑了(51)的如下推广

① 见华罗庚：数论导引，第七章。——译注

$$\sum_{n=1}^{\infty}\frac{\chi(n)}{n^s}. \tag{52}$$

当 $\chi$ 是主特征，即对所有 $(n,m)=1$ 有 $\chi(n)=1$ 时，它有与 $\zeta(s)$ 完全类似的性质，当 $\chi$ 为非主特征时，它可以解析开拓成全复平面上的解析函数，记为 $L(s,\chi)$，称为 Dirichlet $L$-函数。同样可证 $L(s,\chi)$ 在条域 $0\leqslant \mathrm{Re}s\leqslant 1$ 中有无穷多个零点，记为 $\rho_n=\beta_n+i\gamma_n$，则 GRH 可表述为：对模 $m$ 的任何特征，皆有 $\beta_n=1/2$。显然 Riemann 猜想是它的一个特例。这些猜想至今仍未被证明。现在已有的数据表明这些猜想极有可能是正确的。

1976年，G.L.Miller 在一篇短文中证明了关于 Legendre 符号的 Euler 判别法（见本文(10)式）之逆命题成立。

引理5 (Miller)　若 $n$ 为奇合数，则必有自然数 $a$，$(a,n)=1$，使

$$a^{(n-1)/2}\not\equiv\left(\frac{a}{n}\right)\pmod n,$$

其中$\left(\frac{a}{n}\right)$为 Jacobi 符号.

证明　设有奇素数 $p$ 及整数 $\alpha>1$ 使 $p^\alpha\|n$。取 $a$ 为 $p^\alpha$的一个原根，且不妨可设 $(a,n)=1$，否则可求 $a_1$适合

$$\begin{cases}a_1\equiv a \pmod{p^\alpha}\\ a_1\equiv 1 \pmod{n/p^\alpha}.\end{cases}$$

若有

$$a^{(n-1)/2}\equiv\left(\frac{a}{n}\right)\pmod n,$$

则可证有

$$a^{n-1}\equiv 1\pmod{p^\alpha},$$

从而 $\varphi(p^{\alpha})\mid(n-1)$，这与 $p^{\alpha-1}\mid n$ 相矛盾。故必有

$$a^{(n-1)/2}\not\equiv\left(\frac{a}{n}\right)\pmod n.$$

再设 $n=p_1\cdots p_s$，$s\geqslant 2$，诸 $p_i$ 为互不相同的奇素数。取 $b_1$ 为 $p_1$ 的二次非剩余，取 $b_i$ 为 $p_i\ (i=2,\cdots,s)$ 的二次剩余，由孙子定理可求得 $a_1$ 适合

$$\begin{cases}a_1\equiv b_1 \pmod{p_1}\\ a_1\equiv b_i \pmod{p_i},\quad i=2,\cdots,s,\end{cases}$$

则 $(a_1,n)=1$ 且 $\left(\frac{a_1}{n}\right)=-1$。下面用反证法。设若对每个满足 $(b,n)=1$ 的 $b$ 皆有 $b^{(n-1)/2}\equiv\left(\frac{b}{n}\right)\pmod n$，则有 $b^{n-1}\equiv 1\pmod n$，取 $k$ 为 $p_2$ 的原根，则可不妨设 $(k,n)=1$。于是取 $b=k$ 有

$$k^{n-1}\equiv 1\pmod{p_2},$$

从而 $(p_2-1)\mid(n-1)$，即 $\frac{n-1}{p_2-1}$ 是整数。注意到 $a_1$ 是 $p_2$ 的二次剩余，我们就有

$$a_1^{\frac{n-1}{2}}=\left(a_1^{\frac{p_2-1}{2}}\right)^{\frac{n-1}{p_2-1}}\equiv\left(\frac{a_1}{p_2}\right)^{\frac{n-1}{p_2-1}}=1\pmod{p_2},$$

又有

$$a_1^{\frac{n-1}{2}}\equiv\left(\frac{a_1}{n}\right)=-1\pmod n,$$

注意到 $p_2\mid n$ 得 $1\equiv-1\pmod{p_2}$，这与 $p_2>2$ 矛盾。

上述引理只给出了当 $n$ 为奇合数时存在 $a$ 使

$$a^{\frac{n-1}{2}} \not\equiv \left(\frac{a}{n}\right),$$

却并未给出求 $a$ 的方法，也没有给出这种 $a$ 可能存在的范围。

1952年，在假设 GRH 成立的条件下，N.C.Ankeny 对最小二次非剩余的上界给出一个很好的估计，有关这个结论更一般的表述及证明可参见 H.L.Montgomery[35] p.120。利用这些结论，G.L.Miller[34] 给出了判别素性的一种多项式算法，这个算法的一种简化形式如下。

根据 Ankeny-Montgomery 的定理，在假设 GRH 成立的条件下，存在一个整数 $c\geqslant 1$。设 $n$ 是一个待验的奇数。这样可定义函数 $f(n)=c(\ln n)^2$。

第一步　检查是否有某个整数 $m>1$ 及整数 $s\geqslant 2$ 使 $n=m^s$。若这样的 $m$、$s$ 存在，则算法停止并输出"$n$ 是合数"。

第二步　对每个正整数 $a\leqslant f(n)$，作以下三步检查，在任一步成立时，即停止并输出"$n$ 是合数"：

(i) $a|n$,

(ii) $a^{n-1}\not\equiv 1 \pmod n$,

(iii) 对某个 $k$，$1\leqslant k\leqslant t(n-1)$，有

$$\mathrm{GCD}((a^{(n-1)/2^k} \bmod n)-1, n)\neq 1,$$

其中 $t(r)$ 表示 $r$ 中含因子 2 的最高幂次，即 $2^{t(r)}|r$，但 $2^{t(r)+1}\nmid r$。

如果 $n$ 通过以上检查，则输出"$n$ 是素数"。

正是 Miller 的这项成果为强伪素数检测法及 Rabin 的合数检测法提供了富有成效的思想。

习题12　设 $m$ 与 $k$ 为正整数。证明，若从 $[1,m]$ 中独立

且一致地随机选取 $k$ 个整数 $b_1, \cdots, b_k$，则至少有一个奇合数 $n \in [1, m]$ 对所有基 $b_1, \cdots, b_k$ 均为伪素数的概率小于 $m2^{-k-1}$。特别地，有一个由至多 $\ln m/\ln 2$ 个整数组成的集合 $T \subseteq [1, m]$，使得每一个奇合数 $n \in [1, m]$ 对至少一个 $b \in T$ 不满足(47)式。(请与[2]比较之。可惜的是，这个结果并未告诉我们怎样求出一个这样的集合 $T$.)

如果放弃未经证明的 GRH 而改用[9]中的结果，可以得到一个判别素数的无条件算法，不过相应应取算法中的函数为 $f(n) = c_1 n^{1/7}$。详见文献[34]。

(内容未完，待续，请见第(3)册)

(张明尧编译，潘承彪校)

# 洛仑兹几何中的三角学①

G.S.BIRMAN, K.NOMIZU

欧氏几何是仿射空间中在它的向量之间存在正定的内积而产生的几何学.如果把这种内积换成符号为(+,…,+,-)的内积，则所得到的几何就称为洛仑兹几何。如所周知，4维的洛仑兹几何是特殊相对论的最适当的语言。但是，洛仑兹几何的最初等的部分，即洛仑兹平面几何。尚需普及成为数学常识的组成部分。特别是洛仑兹几何中的三角学，即关于洛仑兹空间中三角形的边长和夹角的学问，至今所知甚少。

在本文,我们从所谓的反向的 Cauchy-Schwarz 不等式出发，给出洛仑兹三角学的一个简短的引论。

## 1. 准备知识

设 $L^2$ 是向量空间 $\boldsymbol{R}^2$，并在其中有洛仑兹内积

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = x_1 y_1 - x_2 y_2, \tag{1}$$

其中 $\boldsymbol{x}=(x_1,x_2)$,$\boldsymbol{y}=(y_1,y_2)$。用 $G$ 表示正常的洛仑兹群 $SO^+(1,1)$，它的元素是如下的矩阵：

---

① Trigonometry in Lorentzian Geometry, *Amer. Math. Monthly*, 91(1984), 543—549.

$$A(u)=\begin{bmatrix}\operatorname{ch}u & \operatorname{sh}u\\ \operatorname{sh}u & \operatorname{ch}u\end{bmatrix},\qquad u\in\boldsymbol{R},\qquad(2)$$

其中 ch,sh 分别是双曲余弦函数和双曲正弦函数。与(2)相对照的是转动矩阵

$$R(u)=\begin{bmatrix}\cos u & \sin u\\ \sin u & \cos u\end{bmatrix},\qquad u\in\boldsymbol{R},$$

它保持欧氏内积$\langle\langle \boldsymbol{x},\boldsymbol{y}\rangle\rangle=x_1y_1+x_2y_2$不变。

设 $\boldsymbol{x}$ 是 $L^2$ 中一个向量。如果$\langle\boldsymbol{x},\boldsymbol{x}\rangle<0$，则称 $\boldsymbol{x}$ 是类时向量；如果$\langle\boldsymbol{x},\boldsymbol{x}\rangle>0$，则称 $\boldsymbol{x}$ 是类空向量；如果$\langle\boldsymbol{x},\boldsymbol{x}\rangle=0$，则称 $\boldsymbol{x}$ 是零化向量。$\boldsymbol{x}$ 的模定义为$\|\boldsymbol{x}\|=\sqrt{|\langle\boldsymbol{x},\boldsymbol{x}\rangle|}$.我们还规定时间定向如下：设 $\boldsymbol{e}=(0,1)$，$\boldsymbol{x}=(x_1,x_2)$是类时向量，如果$\langle\boldsymbol{x},\boldsymbol{e}\rangle<0$,则称 $\boldsymbol{x}$ 是指向未来的；如果$\langle\boldsymbol{x},\boldsymbol{e}\rangle>0$，则称 $\boldsymbol{x}$是指向过去的。因此向量 $\boldsymbol{x}=(x_1,x_2)$ 是指向未来的类时向量,当且仅当 $x_1^2-x_2^2<0,x_2>0$；这个条件可以表示成$|x_1|<x_2$。事实上,容易证明：群 $G$ 是 $L^2$ 上保持内积、保持定向、保持时间定向的所有线性变换构成的群。

下面的引理是十分基本的。

引理　设 $\boldsymbol{x},\boldsymbol{y}$ 是 $L^2$ 中指向未来的类时向量，则

(i) $\langle\boldsymbol{x},\boldsymbol{y}\rangle\leqslant 0$。

(ii) $\boldsymbol{x}+\boldsymbol{y}$ 仍是指向未来的类时向量.

(iii) $-\langle\boldsymbol{x},\boldsymbol{y}\rangle\geqslant\|\boldsymbol{x}\|\|\boldsymbol{y}\|$；等号成立当且仅当对于某个$c<0$有 $\boldsymbol{y}=c\boldsymbol{x}$。

(iv) $\|\boldsymbol{x}+\boldsymbol{y}\|\geqslant\|\boldsymbol{x}\|+\|\boldsymbol{y}\|$；等号成立当且仅当对于某个$c>0$ 有 $\boldsymbol{y}=c\boldsymbol{x}$。

如果注意到 $\boldsymbol{x}=(x_1,x_2)$是指向未来的类时向量的充分必

要条件是$|x_1|<x_2$，则引理是很容易直接证明的。另外，引理在任意维的洛仑兹向量空间中都成立，(iii)就是所谓的反向 Cauchy-Schwarz 不等式。向量$\boldsymbol{x}=(x_1,\cdots,x_n,x_{n+1})$和$\boldsymbol{y}=(y_1,\cdots,y_n,y_{n+1})$的洛仑兹内积是

$$<\boldsymbol{x},\boldsymbol{y}>=\sum_{k=1}^{n}x_ky_k-x_{n+1}y_{n+1}.$$

由于$\boldsymbol{x}$是类时指向未来向量，故有$x_{n+1}>\sqrt{\sum_{k=1}^{n}x_k^2}$；同理有$y_{n+1}>\sqrt{\sum_{k=1}^{n}y_k^2}$。利用普通的 Cauchy-Schwarz 不等式

$$\left|\sum_{k=1}^{n}x_ky_k\right|\leqslant\sqrt{\sum_{k=1}^{n}x_k^2}\sqrt{\sum_{k=1}^{n}y_k^2},$$

我们有

$$\begin{aligned}-<\boldsymbol{x},\boldsymbol{y}>&=x_{n+1}y_{n+1}-\sum_{k=1}^{n}x_ky_k\\&\geqslant x_{n+1}y_{n+1}-\sqrt{\sum_{k=1}^{n}x_k^2}\cdot\sqrt{\sum_{k=1}^{n}y_k^2}>0,\end{aligned}$$

因此

$$\begin{aligned}&(-<\boldsymbol{x},\boldsymbol{y}>)^2-\|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2\\&\geqslant\left(x_{n+1}y_{n+1}-\sqrt{\sum_{k=1}^{n}x_k^2}\sqrt{\sum_{k=1}^{n}y_k^2}\right)^2\\&\quad-\left(x_{n+1}^2-\sum_{k=1}^{n}x_k^2\right)\cdot\left(y_{n+1}^2-\sum_{k=1}^{n}y_k^2\right)\\&=x_{n+1}^2\sum_{k=1}^{n}y_k^2-2x_{n+1}y_{n+1}\sqrt{\sum_{k=1}^{n}x_k^2}\sqrt{\sum_{k=1}^{n}y_k^2}\end{aligned}$$

$$+y_{n+1}^2\sum_{k=1}^{n}x_k^2$$

$$=\left(x_{n+1}\sqrt{\sum_{k=1}^{n}y_k^2}-y_{n+1}\sqrt{\sum_{k=1}^{n}x_k^2}\right)^2\geqslant 0,$$

即

$$-\langle \boldsymbol{x},\boldsymbol{y}\rangle\geqslant\|\boldsymbol{x}\|\|\boldsymbol{y}\|.$$

等号成立，当且仅当

$$\sum_{k=1}^{n}x_ky_k=\sqrt{\sum_{k=1}^{n}x_k^2}\cdot\sqrt{\sum_{k=1}^{n}y_k^2},$$

$$x_{n+1}\sqrt{\sum_{k=1}^{n}y_k^2}-y_{n+1}\sqrt{\sum_{k=1}^{n}x_k^2}=0.$$

上面两式成立的条件是存在 $c>0$，使得 $\boldsymbol{y}=c\cdot\boldsymbol{x}$。(iii)式证毕。(iv)是(iii)的直接推论。

现在定义角的概念。设 $\boldsymbol{x},\boldsymbol{y}$ 是两个指向未来的类时单位向量。我们称 $u\in\boldsymbol{R}$ 是从 $\boldsymbol{x}$ 到 $\boldsymbol{y}$ 的(有向)角，如果 $A(u)\cdot\boldsymbol{x}=\boldsymbol{y}$。此时，从 $\boldsymbol{y}$ 到 $\boldsymbol{x}$ 的(有向)角显然是 $-u$。$\boldsymbol{x}$ 和 $\boldsymbol{y}$ 之间的(非有向)角定义为 $|u|$。根据定义，

$$y_1=x_1\operatorname{ch}u+x_2\operatorname{sh}u,$$
$$y_2=x_1\operatorname{sh}u+x_2\operatorname{ch}u,$$

因此

$$-\langle \boldsymbol{x},\boldsymbol{y}\rangle=-y_1x_1+y_2x_2=\operatorname{ch}u. \tag{3}$$

由引理可知，左边的值大于1，所以这样的 $u\in\boldsymbol{R}$ 是完全确定的。

若 $\boldsymbol{x},\boldsymbol{y}$ 是指向未来的类时向量，$\boldsymbol{x}$ 和 $\boldsymbol{y}$ 之间的(有向或非有向的)角就是 $\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}$ 和 $\frac{\boldsymbol{y}}{\|\boldsymbol{y}\|}$ 之间的角。因此

$$\operatorname{ch} u = -\frac{\langle x, y\rangle}{\|x\| \cdot \|y\|}. \tag{4}$$

两条类时直线（它们都通过原点，即通过零向量）之间的类角定义为分别落在这两条直线上的两个类时指向未来的单位向量之间的夹角。

**注记** 两条类空直线之间的夹角也可以类似地定义。其实只要考虑 $L^2$ 上的线性变换 $\sigma$，使得

$$\sigma \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_2 \\ x_1 \end{pmatrix},$$

则对于任意的 $\boldsymbol{x}, \boldsymbol{y} \in L^2$ 有

$$\langle \sigma(\boldsymbol{x}), \sigma(\boldsymbol{y})\rangle = -\langle \boldsymbol{x}, \boldsymbol{y}\rangle,$$

所以 $\sigma$ 把类空向量变为类时向量，把类时向量变为类空向量。若 $\boldsymbol{x}, \boldsymbol{y}$ 是 $L^2$ 上两个指向右① 的类空向量，则 $\sigma(\boldsymbol{x}), \sigma(\boldsymbol{y})$ 是两个指向未来的类时向量，于是由引理得到

(i)′ $\langle \boldsymbol{x}, \boldsymbol{y}\rangle \geqslant 0$,

(iii)′ $\langle \boldsymbol{x}, \boldsymbol{y}\rangle \geqslant \|\boldsymbol{x}\| \|\boldsymbol{y}\|$.

若从 $\sigma(\boldsymbol{x})$ 和 $\sigma(\boldsymbol{y})$ 的角是 $u$，即

$$A(u) \cdot \sigma(\boldsymbol{x}) = \sigma(\boldsymbol{y}),$$

则不难证明

$$A(u) \cdot \boldsymbol{x} = \boldsymbol{y}.$$

所以两个指向未来的类时向量夹角的定义对于两个指向右的类空向量也是适用的，而且变换 $\sigma$ 保持夹角和向量长度不变。

---

① 命 $\boldsymbol{e}_1 = (1, 0)$，则 $\boldsymbol{x}$ 是指向右的类空向量的条件是：$\boldsymbol{x}$ 是类空向量，并且 $\langle \boldsymbol{x}, \boldsymbol{e}_1\rangle > 0$。若 $\langle \boldsymbol{x}, \boldsymbol{e}_1\rangle < 0$，则称 $\boldsymbol{x}$ 是指向左的类空向量。

正因为如此，我们在下面只讨论类时向量。

现在我们转向洛仑兹平面，即与洛仑兹向量空间伴随的仿射平面。这样，对于平面上任意两个点 $A,B$，我们有向量 $\overrightarrow{AB}\in L^2$；对于任意三个点 $A,B,C$，则有 $\overrightarrow{AC}=\overrightarrow{AB}+\overrightarrow{BC}$。

我们知道在欧氏平面上任意一个运动都可以表示成旋转和平移的组合。在直角坐标系下，一个运动用矩阵运算表示就是

$$\begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix} \to \begin{pmatrix} \cos t & -\sin t & a \\ \sin t & \cos t & b \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix}.$$

在洛仑兹平面上，起运动群作用的群是

$$\tilde{G}=\left\{\begin{pmatrix} \operatorname{ch} u & \operatorname{sh} u & a \\ \operatorname{sh} u & \operatorname{ch} u & b \\ 0 & 0 & 1 \end{pmatrix};\ u,a,b\in \boldsymbol{R}\right\}, \tag{5}$$

它在洛仑兹平面上作用的方式与欧氏平面相仿。因此，洛仑兹平面几何就是研究在群 $\tilde{G}$ 的作用下的不变性质。

设 $A,B,C$ 是不共线的三个点，使得 $\overrightarrow{AB},\overrightarrow{AC}$ 是指向未来的类时向量，$\overrightarrow{BC}$ 是类空向量，并且 $\langle\overrightarrow{AB},\overrightarrow{BC}\rangle=0$，则 $B$ 就是从点 $C$ 向直线 $AB$ 作正交投影得到的。因此

$$\begin{aligned}
\langle\overrightarrow{AB},\overrightarrow{BC}\rangle &= \langle\overrightarrow{AB},\overrightarrow{AC}-\overrightarrow{AB}\rangle \\
&= \langle\overrightarrow{AB},\overrightarrow{AC}\rangle-\langle\overrightarrow{AB},\overrightarrow{AB}\rangle \\
&= -\|\overrightarrow{AB}\|\cdot\|\overrightarrow{AC}\|\cdot\operatorname{ch} u+\|\overrightarrow{AB}\|^2 \\
&= 0,
\end{aligned}$$

即直线 $AB$ 和 $AC$ 之间的角 $u$ 满足

$$\operatorname{ch} u=\frac{\|\overrightarrow{AB}\|}{\|\overrightarrow{AC}\|}. \tag{6}$$

另外，

$$\begin{aligned}
\langle\overrightarrow{BC},\overrightarrow{BC}\rangle &= \langle\overrightarrow{AC}-\overrightarrow{AB},\overrightarrow{BC}\rangle \\
&= \langle\overrightarrow{AC},\overrightarrow{AC}\rangle-\langle\overrightarrow{AC},\overrightarrow{AB}\rangle \\
&= -\|\overrightarrow{AC}\|^2+\|\overrightarrow{AB}\|\cdot\|\overrightarrow{AC}\|\cdot\operatorname{ch} u \\
&= \|\overrightarrow{AC}\|^2(\operatorname{ch}^2 u-1)=\|\overrightarrow{AC}\|^2\operatorname{sh}^2 u,
\end{aligned}$$

因此

$$\operatorname{sh} u=\frac{\|\overrightarrow{BC}\|}{\|\overrightarrow{AC}\|}, \tag{7}$$

并且

$$\|\overrightarrow{AC}\|^2+\|\overrightarrow{BC}\|^2=\|\overrightarrow{AB}\|^2. \tag{8}$$

现在我们引进在洛仑兹平面几何中很重要的一类三角形。所谓(类时的)纯三角形是指有三个适当命名的顶点$A,B,C$的三角形，使得$\overrightarrow{AB},\overrightarrow{BC}$都是指向未来的类时向量(作为引理的推论可知，$\overrightarrow{AC}$也是指向未来的类时向量)。以下我们假定纯三角形$ABC$的顶点都是这样命名的，$B$称为中间顶点。在$A$处的角$\hat{A}$是直线$AB$和$AC$之间的角；在$C$处的角$\hat{C}$是直线$BC$和$AC$之间的角。最后，在中间顶点$B$处的角$\hat{B}$是直线$AB$和$BC$之间的角，也就是向量$\overrightarrow{AB}$和$\overrightarrow{BC}$之间的角，它看上去象是欧氏几何中三角形的外角。

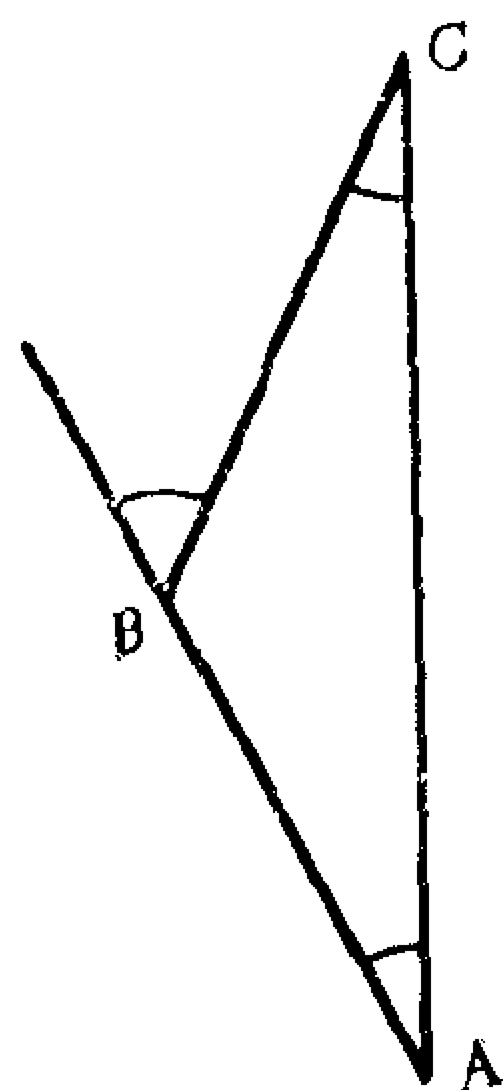


图 1

参看图1。

现在我们证明

**命题1** 设 $ABC$ 是纯三角形，则在中间顶点的角等于另外两个角之和：$\hat{B}=\hat{A}+\hat{C}$。

**证明** 设 $\boldsymbol{x}=\overrightarrow{AB}/\|\overrightarrow{AB}\|$，$\boldsymbol{y}=\overrightarrow{BC}/\|\overrightarrow{BC}\|$，$\boldsymbol{z}=\overrightarrow{AC}/\|\overrightarrow{AC}\|$，设 $\alpha$ 是从 $\boldsymbol{x}$ 到 $\boldsymbol{z}$ 的角，$\beta$ 是从 $\boldsymbol{x}$ 到 $\boldsymbol{y}$ 的角，$\gamma$ 是从 $\boldsymbol{z}$ 到 $\boldsymbol{y}$ 的角。因为 $\boldsymbol{z}$ 落在 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 之间(更确切地说,若 $\boldsymbol{x}=(\mathrm{sh}\,u,\mathrm{ch}\,u)$，$\boldsymbol{y}=(\mathrm{sh}\,v,\mathrm{ch}\,v)$，$\boldsymbol{z}=(\mathrm{sh}\,w,\mathrm{ch}\,w)$，则 $w$ 落在 $u$ 和 $v$ 之间)，显然有 $\beta=\alpha+\gamma$。因为 $\alpha$ 和 $\gamma$ 同时是正的，或同时是负的，因此 $|\beta|=|\alpha|+|\gamma|$。

在洛仑兹平面几何中平行四边形面积的概念是有意义的。给定两个向量 $\boldsymbol{x}=(x_1,x_2)$，$\boldsymbol{y}=(y_1,y_2)$，则由 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 张成的平行四边形的面积等于 $|x_1y_2-x_2y_1|$，它在群 $G$ 的作用下是不变的(只要注意到 $G$ 是么模群 $\mathrm{SL}(2,\boldsymbol{R})$ 的子群)。如果 $A$，$B$，$C$ 是不共线的三个点，则三角形 $ABC$ 的面积等于由 $\overrightarrow{AB}$，$\overrightarrow{AC}$ 所张的平行四边形的面积的一半，它在群 $\tilde{G}$ 的作用下是不变的。我们有

**命题2** 纯三角形 $ABC$ 的面积为

$$S=\frac{1}{2}bc\,\mathrm{sh}\,\hat{A}, \tag{9}$$

其中 $b=\|\overrightarrow{AC}\|$，$c=\|\overrightarrow{AB}\|$，$\hat{A}$ 是 $\triangle ABC$ 在 $A$ 处的角(图2)。

**证明** 不妨设 $A=(0,0)$，$B=c\cdot(\mathrm{sh}\,u,\mathrm{ch}\,u)$，$C=b\cdot(\mathrm{sh}\,v,\mathrm{ch}\,v)$，则纯三角形 $ABC$ 的面积是

$$S=\frac{1}{2}bc|\mathrm{sh}\,u\,\mathrm{ch}\,v-\mathrm{ch}\,u\,\mathrm{sh}\,v|=\frac{1}{2}bc\,\mathrm{sh}|u-v|,$$

很明显，$|u-v|$ 恰好等于角 $\hat{A}$。

## 2. 洛仑兹圆周

洛仑兹平面上的圆周是如下定义的：设$P$是洛仑兹平面上任意一点，设$r>0$。曲线$(Q: \langle \overrightarrow{PQ}, \overrightarrow{PQ} \rangle = r^2)$ 将有两个分支，每一个分支称为以$P$为中心，以$r$为半径的（类时）洛仑兹圆周。设$P=(p_1,p_2)$，则洛仑兹圆周的方程是

$$(x_1-p_1)^2-(x_2-p_2)^2=r^2.$$

从微分几何观点看，这条双曲线的每一个分支都能看作一条有常曲率的类时曲线（这里的"类时"是指它的切向量是类时向量）。事实上，一个分支可以参数化为$\boldsymbol{x}(t)=(r\,\text{ch}\,t+p_1,\ r\,\text{sh}\,t+p_2)$，其切向量是$\frac{d\boldsymbol{x}}{dt}=(r\,\text{sh}\,t, r\,\text{ch}\,t)$，所以

$$\left\langle \frac{d\boldsymbol{x}}{dt}, \frac{d\boldsymbol{x}}{dt} \right\rangle = -r^2.$$

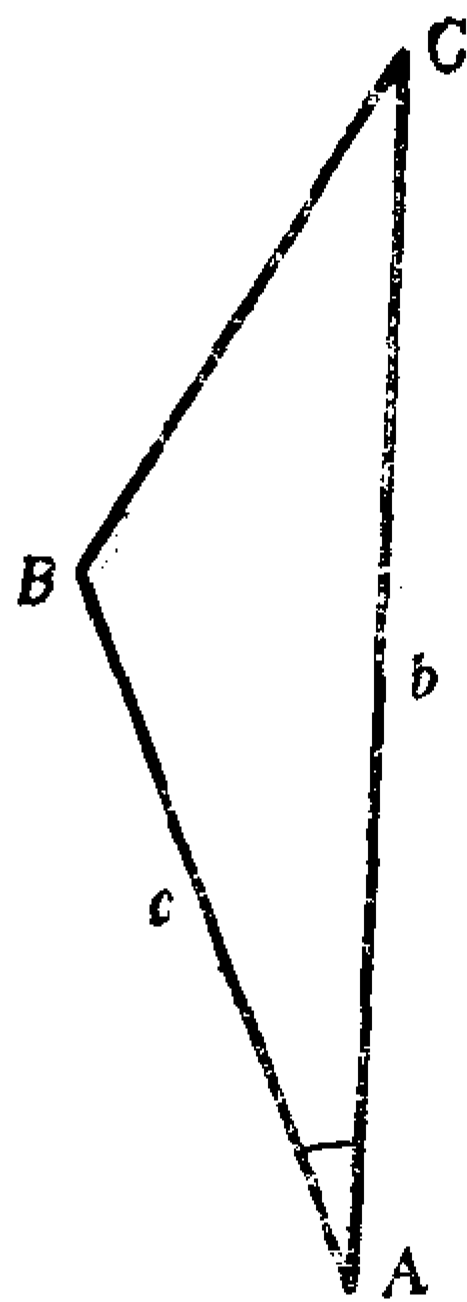


图 2

容易看出在洛仑兹圆上任意一点$Q$，$\overrightarrow{PQ}$与洛仑兹圆在$Q$处的切线$T_Q$关于洛仑兹内积是垂直的，即$\overrightarrow{PQ}$是洛仑兹圆周在$Q$点的法向量。在欧氏几何中相应的结果也成立，即圆周的半径向量与该点处的切线是彼此垂直的。这样，洛仑兹圆周的单位法向量是$\boldsymbol{n}=(\text{ch}\,t,\text{sh}\,t)$，曲率定义为$\kappa=\left\langle \frac{d^2\boldsymbol{x}}{ds^2}, \boldsymbol{n} \right\rangle$，其中$s$是弧长参数。显然$\frac{ds}{dt}=\left\|\frac{d\boldsymbol{x}}{dt}\right\|=r$，所以$\frac{d^2\boldsymbol{x}}{ds^2}=\frac{1}{r}(\text{ch}\,t,\text{sh}\,t)$，$\kappa=\frac{1}{r}$。因此，洛仑兹圆周有常曲率

$\frac{1}{r}$.

现在我们有

**定理3** 洛仑兹圆周的内接三角形是纯三角形。反之，一个纯三角形必能内接于唯一的一个洛仑兹圆周。

**证明** 设 $A = r\cdot(\text{ch}\, a, \text{sh}\, a)$，$B = r\cdot(\text{ch}\, b, \text{sh}\, b)$，$C = r\cdot(\text{ch}\, c, \text{sh}\, c)$是洛仑兹圆周 $x^2 - y^2 = r^2$ 上的三个点，其中 $a < b < c$。由于 $a \neq b$，故有

$$\langle \overrightarrow{AB}, \overrightarrow{AB} \rangle = r^2[(\text{ch}\, b - \text{ch}\, a)^2 - (\text{sh}\, b - \text{sh}\, a)^2]$$
$$= 2r^2[1 - \text{ch}(b-a)] < 0,$$

所以$\overrightarrow{AB}$是类时向量。容易验证 $\overrightarrow{AB}$ 是指向未来的。同理可证$\overrightarrow{BC}$是指向未来的类时向量，因此 $ABC$ 是纯三角形。

现在设 $ABC$ 是纯三角形，命 $A = (a_1, a_2)$，$B = (b_1, b_2)$，$C = (c_1, c_2)$。我们要证明存在 $p_1, p_2$，及 $r \neq 0$，使得下式成立：

$$\begin{cases} (a_1 - p_1)^2 - (a_2 - p_2)^2 = r^2, \\ (b_1 - p_1)^2 - (b_2 - p_2)^2 = r^2, \\ (c_1 - p_1)^2 - (c_2 - p_2)^2 = r^2. \end{cases} \tag{10}$$

从（10）式我们得到 $p_1, p_2$ 的两个线性方程

$$\begin{cases} 2(b_1 - a_1)p_1 + 2(a_2 - b_2)p_2 = b_1^2 - a_1^2 + a_2^2 - b_2^2, \\ 2(c_1 - b_1)p_1 + 2(b_2 - c_2)p_2 = c_1^2 - b_1^2 + b_2^2 - c_2^2. \end{cases} \tag{11}$$

因为$\overrightarrow{AB}$与 $\overrightarrow{BC}$ 不共线，于是 $(b_1 - a_1)\cdot(b_2 - c_2) \neq (a_2 - b_2)\cdot(c_1 - b_1)$，因此方程(11)有唯一解 $p_1, p_2$。现在从（10)的第一个方程决定 $r$，则其余两个方程也满足。剩下要验证的是 $r \neq 0$。

若设 $r=0$，则我们必定有

$$a_1-p_1=\pm(a_2-p_2),$$
$$b_1-p_1=\pm(b_2-p_2),$$
$$c_1-p_1=\pm(c_2-p_2).$$

其中至少有两个正号或有两个负号。如果 $a_1-p_1=a_2-p_2$，$b_1-p_1=b_2-p_2$，则得 $b_1-a_1=b_2-a_2$，这意味着 $\langle\overrightarrow{AB},\overrightarrow{AB}\rangle=0$，这与 $\overrightarrow{AB}$ 是类时向量相矛盾。类似地，其余情况也都不可能成立。因此 $r\neq0$.

我们还得考察洛仑兹圆周所具有的、与欧氏几何的熟知事实类似的进一步的性质。

设 $A,B$ 是洛仑兹圆周上 $x^2-y^2=r^2$，$x>0$ 上的两个点。则 $OA$ 和 $OB$ 都是类空直线。按照(4)式下面的注记，能够定义 $OA$ 和 $OB$ 之间的角，称之为弦 $AB$ 所对的圆心角。如果 $A=r\cdot(\mathrm{ch}\,a,\mathrm{sh}\,a)$，$B=r\cdot(\mathrm{ch}\,b,\mathrm{sh}\,b)$，则弦 $AB$ 所对的圆心角等于 $|a-b|$。

设 $T_B$ 是洛仑兹圆周在 $B$ 点的切线，则我们能够谈论 $T_B$ 和弦 $AB$ 之间的角。我们要证明

**定理4** 设 $A,B$ 是洛仑兹圆周 $x^2-y^2=r^2$，$x>0$ 上的两个点，则切线 $T_B$ 和弦 $AB$ 之间的角等于弦 $AB$ 所对的圆心角的一半（参看图3）。

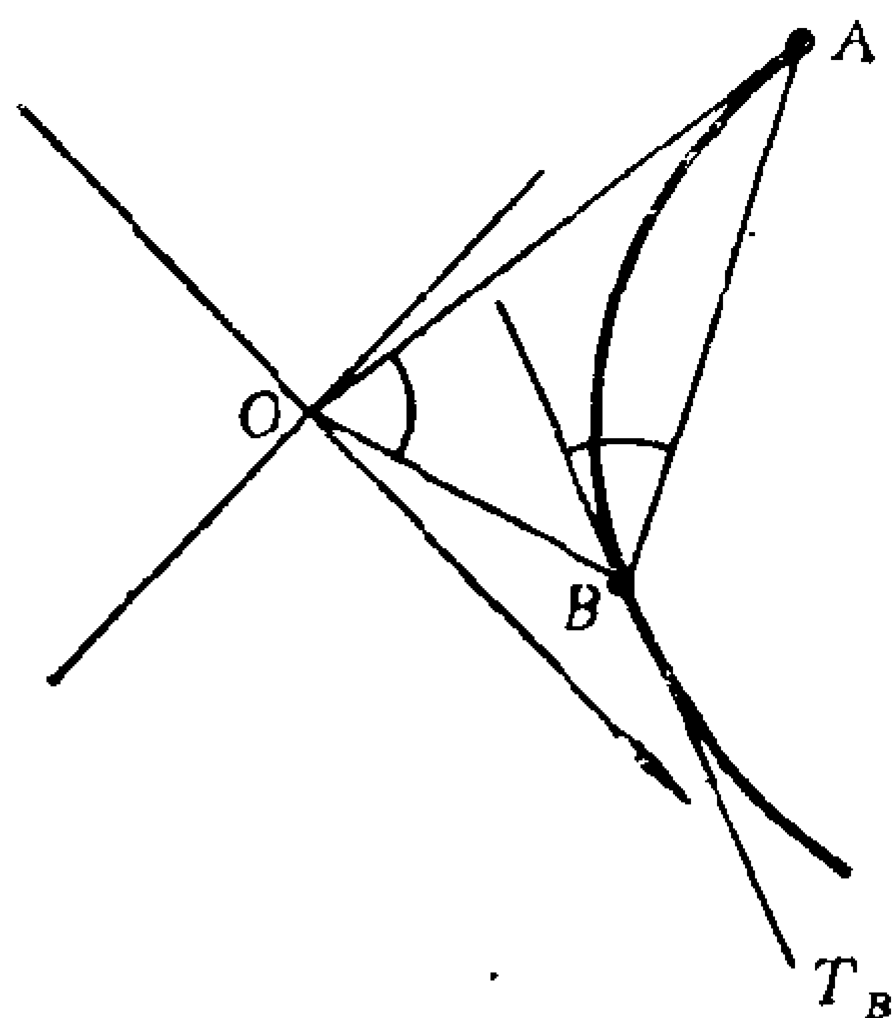


图 3

**证明** 利用群 $G$ 的变换，可以假定 $A=r(\mathrm{ch}\,a,\mathrm{sh}\,a)$，$B=(r,0)$。$T_B$ 和 $AB$ 之间的角 $\alpha$ 满足

$$\operatorname{ch} \alpha = -\langle x, y\rangle / \|y\|,$$

其中 $x=(0,1)$，$y=(\operatorname{ch} a-1, \operatorname{sh} a)$。因而

$$\operatorname{ch} \alpha = \operatorname{sh} a / \sqrt{-(\operatorname{ch} a-1)^2+\operatorname{sh}^2 a}$$
$$= \operatorname{sh} a / \sqrt{2(\operatorname{ch} a-1)}\ .$$

由于 $\operatorname{ch} a-1=2 \operatorname{sh}^2(a/2)$，$\operatorname{sh} a=2\operatorname{sh}(a/2)\operatorname{ch}(a/2)$，我们有 $\operatorname{ch} \alpha=\operatorname{ch}(a/2)$，即 $\alpha=|a|/2$。显然，$|a|$恰好是该弦所对的圆心角。

**推论1** 设$A$和$B$是洛仑兹圆周上的两个点，则弦$AB$和切线$T_A$，$T_B$所构成的角相等（参看图4）。

**推论2** 设$ABC$是内接于洛仑兹圆周的纯三角形。则在中间顶点$B$的角等于弦$AC$所对的圆心角的一半。因此，当点$B$在$A$、$C$之间的洛仑兹圆周上变动时，三角形$ABC$在点$B$的角是定角。

**证明** 参看图5。设$\alpha, \gamma$分别是切线$T_B$与弦$AB$、弦$BC$所成的角。那么在点$B$的角是$\alpha+\gamma$。从定理4知道弦$AB$的圆心角是$2\alpha$，弦$BC$的圆心角是$2\gamma$。这样，弦$AC$所对的圆心角是$2(\alpha+\gamma)$。

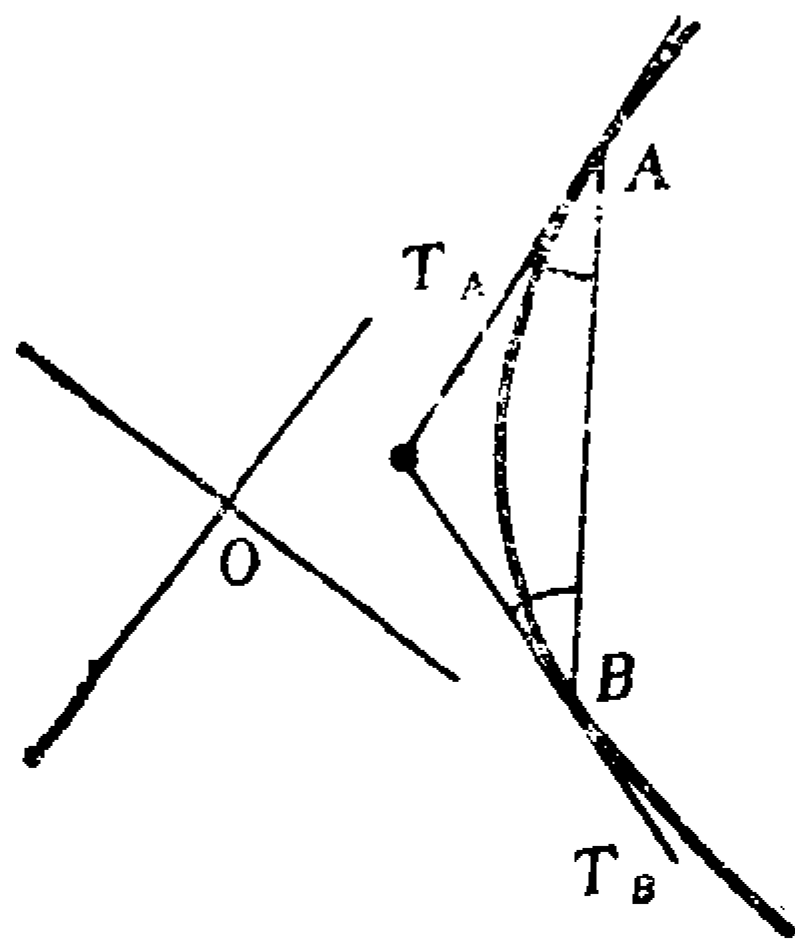



图 4

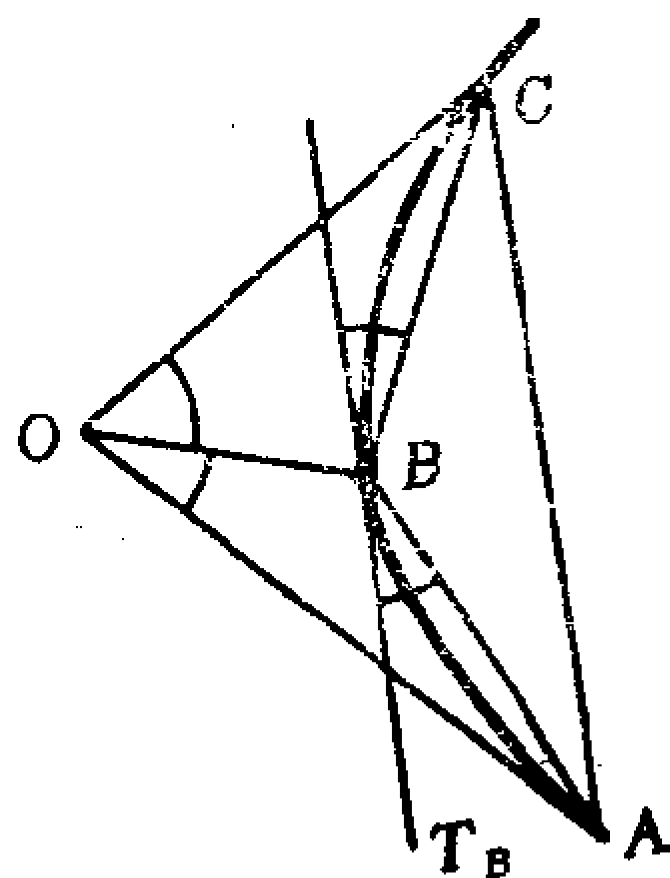



图 5

## 3. 双曲正弦定律和双曲余弦定律

我们要证明

**定理5** 设 $ABC$ 是纯三角形。则

$$\frac{\operatorname{sh}\hat{A}}{a}=\frac{\operatorname{sh}\hat{B}}{b}=\frac{\operatorname{sh}\hat{C}}{c}\quad(\text{双曲正弦定律}),$$

其中 $a=\|\overrightarrow{BC}\|$，$b=\|\overrightarrow{CA}\|$，$c=\|\overrightarrow{AB}\|$，且 $\hat{A},\hat{B},\hat{C}$ 分别记在 $A,B,C$ 的角。

**证明** 参看图6。设 $D$ 是 $AC$ 上一点，使得 $\overrightarrow{BD}$ 是类空向量，并且 $\langle\overrightarrow{BD},\overrightarrow{AC}\rangle=0$。由(7)式得到

$$\operatorname{sh}\hat{A}=\frac{\|\overrightarrow{BD}\|}{c},\qquad \operatorname{sh}\hat{C}=\frac{\|\overrightarrow{BD}\|}{a},$$

因此

$$\frac{\operatorname{sh}\hat{A}}{a}=\frac{\operatorname{sh}\hat{C}}{c}.$$

设 $E$ 是直线 $AB$ 上一点，使得 $\overrightarrow{CE}$ 是类空向量，并且 $\langle\overrightarrow{CE},\overrightarrow{AB}\rangle=0$。这样，

$$\operatorname{sh}\hat{B}=\frac{\|\overrightarrow{CE}\|}{a},\qquad \operatorname{sh}\hat{A}=\frac{\|\overrightarrow{CE}\|}{b},$$

因此

$$\frac{\operatorname{sh}\hat{B}}{b}=\frac{\operatorname{sh}\hat{A}}{a}.$$

**定理6** 定理5中的公比等于 $\frac{1}{2r}$，其中 $r$ 是纯三角形 $ABC$

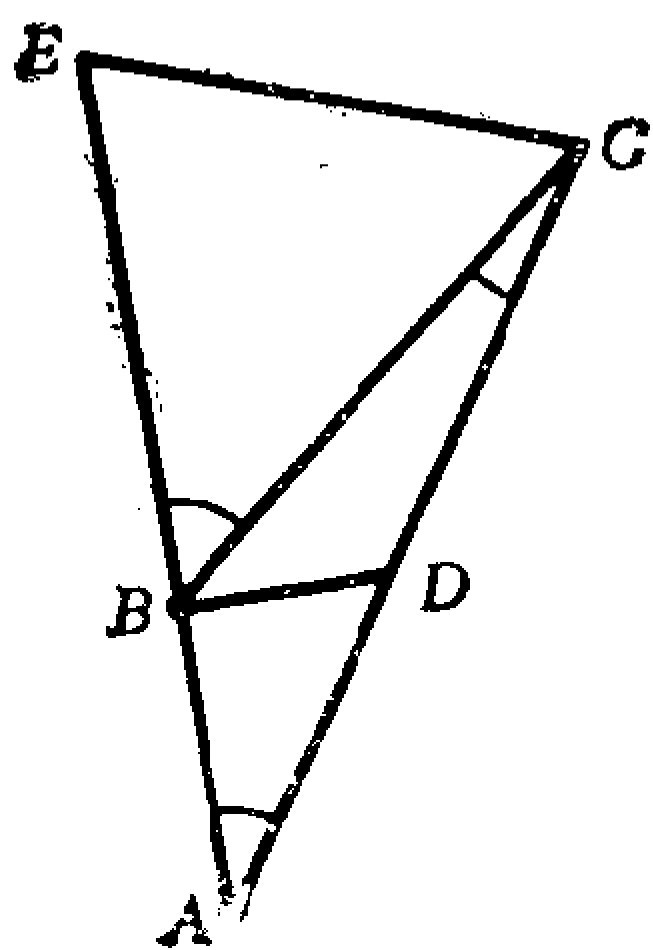


图 6

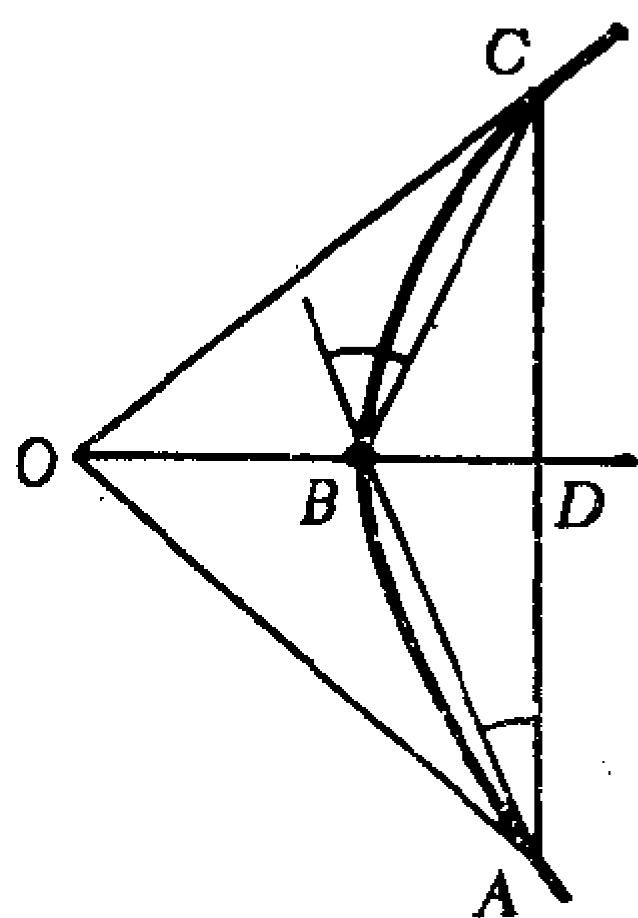


图 7

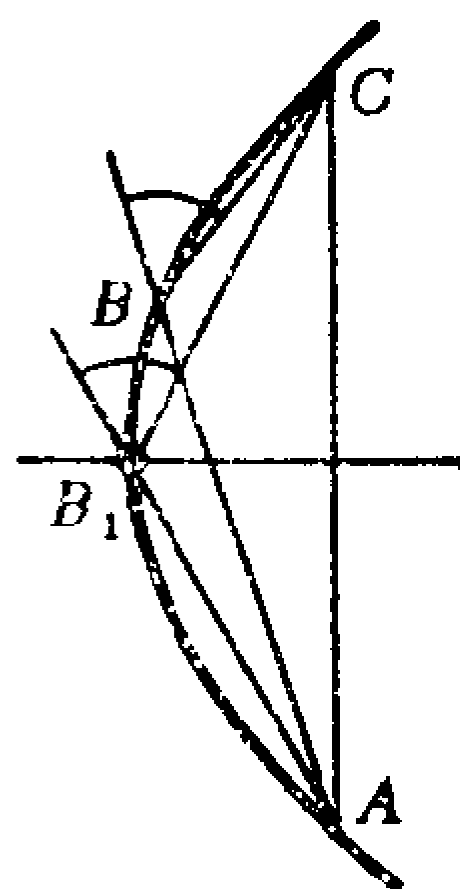


图 8

的外接洛仑兹圆周的半径。

**证明** 首先就$B$落在$x$-轴上、$A$和$C$关于$x$-轴对称的特殊情形证明这个定理（参看图7）。对于$\triangle ABC$，用$\alpha$记相等的角$\hat{A}$和$\hat{C}$。由命题1，我们有$\hat{B}=2\alpha$。由定理4的推论2，弦$BC$的圆心角等于$2\alpha$。如果$D$是$AC$与$x$-轴的交点，则将公式(7)用于$\triangle ODC$给出

$$\|\overrightarrow{CD}\| = r\cdot\mathrm{sh}(2\alpha) = 2r\ \mathrm{ch}\ \alpha\ \mathrm{sh}\ \alpha.$$

将同一个公式用于$\triangle CBD$则给出

$$\|\overrightarrow{CD}\| = \|\overrightarrow{CB}\| \operatorname{ch} \alpha = a \operatorname{ch} \alpha.$$

因此 $2r \operatorname{sh} \alpha = a$，即$\dfrac{\operatorname{sh}\hat{A}}{a} = \dfrac{1}{2r}$.

对于一般情形，先用群$G$的作用使得$A, C$关于$x$-轴是对称的。如果$B_1$是洛仑兹圆周和$x$-轴的交点（看图8），则从定理4的推论2得知$\triangle ABC$的角$\hat{B}$等于$\triangle AB_1C$的角$\hat{B}_1$。这样，$\dfrac{\operatorname{sh}\hat{B}}{\|\overrightarrow{AC}\|} = \dfrac{\operatorname{sh}\hat{B}_1}{\|\overrightarrow{AC}\|}$。根据在特殊情形已证明的事实，后者等于$\dfrac{1}{2r}$，因此$\dfrac{\operatorname{sh}\hat{B}}{b} = \dfrac{1}{2r}$.

**推论** 内接于半径为$r$的洛仑兹圆周的纯三角形$\triangle ABC$的面积等于$\dfrac{abc}{4r}$，也等于$\dfrac{b^2 \operatorname{sh}\hat{A} \operatorname{sh}\hat{C}}{2 \operatorname{sh}\hat{B}}$.

**证明** 这是命题 2 和定理 6 的直接推论。

最后我们有

**定理7** 对于纯三角形$ABC$有

$$a^2 = b^2 + c^2 - 2bc \operatorname{ch} \hat{A},$$

$$c^2 = a^2 + b^2 - 2ab \operatorname{ch} \hat{C}, \quad \text{（双曲余弦定律）}$$

$$b^2 = a^2 + c^2 + 2ac \operatorname{ch} \hat{B}.$$

**证明** 这是容易证明的，留给读者自己证明。需要指出的是，与中间顶点$B$对应的公式中最后一项的符号是$+$号。

## 参 考 文 献


[1] G. S. Birman and K. Nomizu, The Gauss-Bonnet theorem for 2-dimensional spacetimes, *Michigan Math.J.*,

[2] W. H. Greub, Linear Algebra, 2nd ed., Springer-Ver-

lag and Academic Press, New York, 1963.
[3] B. O'Neill, Semi-Riemannian Geometry, Academic Press, 1983.
[4] R. K. Sachs and H. Wu, General Relativity for Mathematicians, Springer-Verlag, New York, 1977.
[5] E. H. Schröder, Gemeinsame Eigenschaften euklidischer, galileischer und minkowskischer Ebenen, Mitteilungen der mathematischen Gesell. schaften Hamburg, X (1974), 185—217.
[6] J. L. Synge, Relativity, The Special Theory, North-Holland and Interscience, 1956.
[7] I. M.Yaglom, A Simple Non-Euclidean Geometry and its Physical Basis, Springer-Verlag, New York, 1979.

（陈维桓编译，阮培文校）

# HANOI塔问题及算法分析①

P. Cull, E.F.Ecklund

## 一、前　言

到处都用到数学，但很可能没有哪一处能比得上它在计算机科学中用得这么广；而在计算机科学中，可能没有哪一处比得上它在算法分析中的应用。如果想划条界限，以此说明“算法分析的这一半属于数学，另一半属于计算机科学”，这可太难了！

本文将通过详细讨论Hanoi塔问题的一些算法给大家一个关于算法分析的感性认识。首先，我们应该感谢唐纳德·克努思先生。他所著的《计算机程序设计技巧》一书[5]是算法分析领域中的权威性著作。近年来，算法分析已经成为计算机科学专业本科生的必修课程。克努思的著作仍然是权威性的参考书，但作为教科书更广泛地采用的是A. 阿赫，J. 赫皮克诺夫和J. 乌尔曼（简记为AHU）合著的《计算机算法的设计与分析》[1]，S. 贝思著的《计算机算法》[2]，以及E. 赫罗维兹和S. 撒尼合著的《计算机算法基础》[4]。

算法分析的任务主要有三方面：

① Tower of Hanoi and Analysis of Algorithms, *Amer. Math. Monthly*, 92(1985), 407—420.

(1) 提出一个可证明其正确性的算法，也就是说，所设计的算法不仅仅能用来解这个问题，而且还要从理论上能够证明这一点。

(2) 对同一个问题提出的若干算法，按它们占用计算机资源（如时间、空间）的各种度量标准进行比较。据此，我们才能说何时一个算法比另一个算法好一些。

(3) 如果可能的话，对给定问题找出一个按占用某种特定资源的度量来说是最好的算法。这包括了证明一个"下界"，即证明解决这个问题的每一个算法都至少需要如此多的这种资源。为了证实一个算法是最好的，人们必须同时有一个下界的证明，以及一个所用资源不超过这个下界的算法。

这里要提醒人们注意的是一个算法的界与一个问题的界之间的区别。如果人们对于解决某个问题的一个算法确定了其占用某种资源的上界，那么对这个算法以及对这个问题来说，我们就都有了一个上界。如果我们确定了某个问题的一个下界，那么也就得到了关于解这个问题所有算法的一个下界。但是，所论证的解某个问题的一个算法的下界，并不能确定这个问题的下界。

在本文中，我们将用Hanoi塔问题作为典型实例来说明算法分析的这三方面任务。Hanoi塔问题常常被当作一个可以用递归算法简捷求解的例子，一个需要指数时间才能求得其解的例子[5]，一个解题的策略的例子[6]。Hanoi塔问题是这样的：给我们三个塔A，B，C和$n$个不同尺寸的圆盘，开始时圆盘按大小顺序摆在A塔上（最大的盘$n$在底部，最小的盘1在顶部，如图1），要求我们将这一摞圆盘从A塔移到C塔，条件是一次只能移动一个盘子且大盘不能摞

在小盘之上。另外，要求移动序列要尽可能的短。一个算法如果解决了河内塔问题，那么当输入圆盘数 $n$ 和塔名时，此算法就产生出符合上述规则的最短移动序列。

本文将研究解 Hanoi 塔问题的种种算法。我们将证明每个算法的正确性，计算出每个算法所用的时间和空间，以便

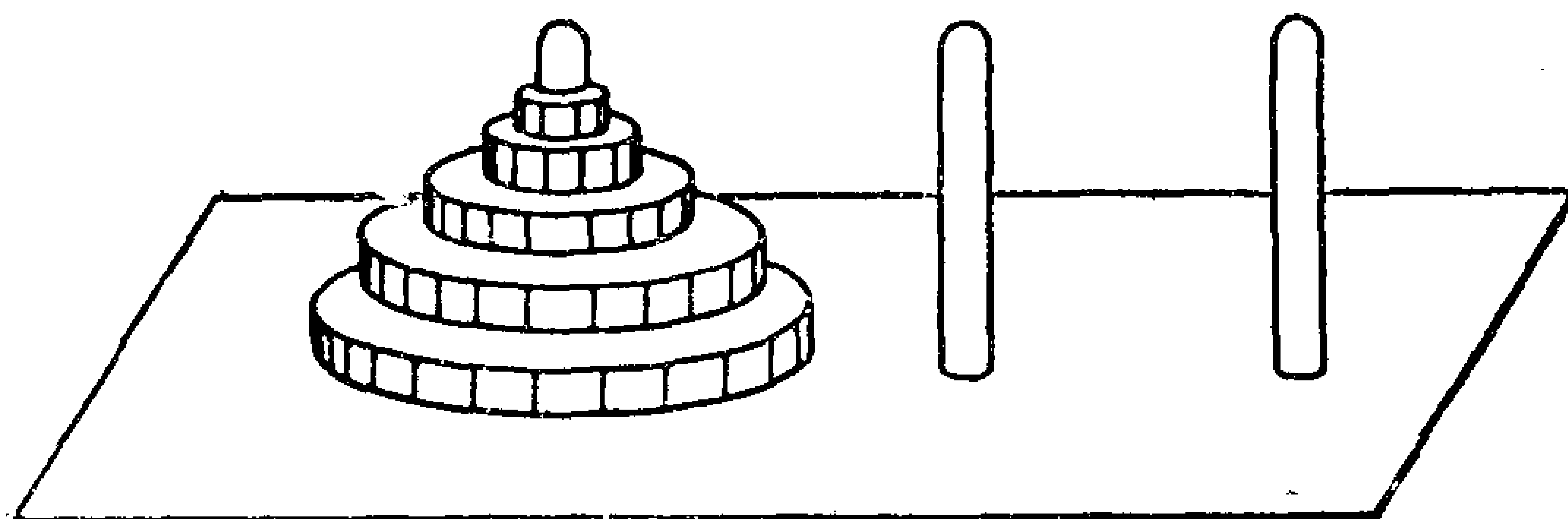


图1 Hanoi 塔问题

于对它们进行比较，并且证明每个算法在解决此问题时所需时间和空间的下界。我们还将说明，最后给出的那个算法达到了这些下界，因此对这些度量来说是最佳可能的。

## 二、算 法 比 较

对任何一个可解的问题都存在无数解此问题的算法。我们怎样才能判定哪一个是最好的算法呢？有几种可能的比较它们的方式。我们将致力于两种度量：时间和空间。如果我们在一台计算机上同时运行两种算法，用较少时间的那个算法将首先结束，我们就说这个算法比另一个算法快。但遗憾的是，为进行这种公平的检验，我们必须保持一些恒定的条件，如必须用相同的程序语言编译这两种算法，必须用相同

的程序编制器编辑这两个程序，并在同台机器上，在同一种操作系统条件下运行这两个程序，且在运行过程中两个程序不产生任何冲突。事实上，即使我们满足了所有这些条件，仍然会懊恼地发现在条件 $C$ 下算法 $A$ 较快，而在条件 $D$ 下算法 $B$ 较快。

为避免发生这种令人不快的情况，我们仅考虑计算时间的阶数。设 n 是问题规模的某种度量，运行时间是 n 的一个函数。例如，在 Hanoi 塔问题中，n 代表圆盘数。对相同阶数的运行时间，我们不加以区分。为此目的，我们把两个 n 的函数 $f(n)$ 和 $g(n)$ 有相同的阶定义为：如果对某个数 $N$，存在两个正的常数 $C_1$ 和 $C_2$，使得对所有的 $n \geqslant N$ 有

$$C_1|g(n)| \leqslant |f(n)| \leqslant C_2|g(n)|.$$

我们把这个关系记为 $f(n)=\theta(g(n))$，读作 $f(n)$ 的阶为 $g(n)$。如果两个算法的运行时间有相同的阶，我们就认为这两个算法所用时间相同。特别地，我们不再区分那些运行时间具有常数倍关系的算法。

如果我们发现算法 $A$ 有一个严格小于算法 $B$ 的时间阶，那么，不管它们的实际运行条件如何，我们可以确信对任何充分大的问题算法 $A$ 比算法 $B$ 运行得快。另一方面，如果算法 $A$ 和 $B$ 有相同的时间阶，那么我们不再探讨在给定的实际条件下，哪一个算法运行的更快些.

一个算法所占用的空间就是这个算法为存储和操作数据所用的毕特（bit，即二进位数，$1\text{bit}=\log_2 2$）个数。[3] 我们认为这个空间是问题规模 $n$ 的一个增函数。测定占用空间时略去说明算法所用的毕特数。因为说明部分有一个与问题规模不相干的大小固定的毕特占用量。由于选择了毕特作为度量空间

的单位，我们可以比计算时间更精确地来计算空间。人们能够将一个占用3n 毕特的算法与一个占用2n 毕特的算法加以区别，但对一个占用3n + 7毕特的算法与一个占用3n + 1毕特的算法就不再加以区别，因为我们可以把常数量的毕特放在算法本身之内。

因此，如果我们能证明某算法具有最小的时间阶和占用最少的空间（顶多相差一个常数），那么就可以说我们有了解决那个问题的最好算法。

但这种最好的算法是否一定存在，我们并不清楚。对某些问题来说，对时间、空间的要求不能两全其美：一个较快的算法需要较多的空间。在 Hanoi 塔问题中不存在这种情况，我们将在最后给出一个同时达到最小时间和最小空间的算法。

## 三、一个递归算法

从某个算法开始，人们不断设法加以改进，以求达到最好的算法。人们常用某些策略来形成一个算法。一个经常用的策略是观察这个问题并且看看它的解是否能化成一些同样的但规模较小的问题。这种策略通常称之谓“分而治之”。如果一个问题可以被分而治之，那么就可以构造一个递归算法来解它。这种构造法几乎同时给出了一个算法正确性的归纳证明。一个分而治之算法的时间和空间分析通常是简单明了的，因为算法直接给出了时间和空间占用量的差分方程。

虽然这些分而治之的算法具有很多好的性质，但是它们可能达不到最小的时-空占用量，然而可以将它们作为构造更有效算法的一个起点。

观察 Hanoi 塔问题，我们发现解决的关键是：移动最大的圆盘前需要先把所有其他的圆盘移出去。因此这n－1个较小的圆盘应该已移到B塔上，但这恰恰是具有较少圆盘的另一个 Hanoi 塔问题。当最大的圆盘由A塔移到C塔后，这n-1个较小的圆盘就可以从B塔移到C塔，而这又是一个较小的 Hanoi 塔问题。这些发现导出了下面的递归算法①。

```
PROCEDURE HANOI(A,B,C,n)
    IF n=1 THEN 将A塔顶端的圆盘移动到C塔上
           ELSE HANOI(A,C,B,n-1)
                将A塔顶端的圆盘移动到C塔上
                HANOI(B,A,C,n-1)。
```

这是一个最好的算法吗？我们将证明这个算法具有最小的时间复杂度，但不具有最小的空间复杂度。首先，我们证明该算法正确地解决了 Hanoi 塔问题。这是在前言中提到的算法分析的第一个任务。

**命题1** 递归算法HANOI正确地解决了河内塔问题。

**证明** 显然，算法正确地给出了只有一个圆盘的最小移动序列。当不只一个圆盘时，该算法把n－1个圆盘移到B塔，然后将最大的圆盘移到C塔，再把这n－1个圆盘从B塔移动到C塔上。这恰好是我们所需要的最少移动算法，因为按照规则，只有当上面n－1个圆盘都移到单独一个塔上时，最大的圆盘才能移动。因此，这n－1个盘必须从A塔移到另一个塔上去。显然，将最大的圆盘从A塔移到C塔，至少需要一次移动。当最大的圆盘移到C塔后，其余的n－1个摞在

---

① 在下面命题1的证明中，对这算法作出了清楚的说明。——译注

B塔上的圆盘还需移到C塔上去。根据归纳假设，这n－1个圆盘按最少移动序列移动，从而对n个圆盘的这个算法不会多做超过最少移动次数的移动，并且当所有n个圆盘都从A塔移到C塔后结束。证毕。

值得注意的是，对此问题我们不仅给出了可证明其正确性的算法，同时还证明了最少移动序列是唯一的。这种唯一性使正确性的证明变得容易了。如果可能存在几种最少移动序列，那么证明将会非常麻烦。

我们希望计算HANOI程序的运行时间，但是尚不知道各种操作需要多长时间。移动一个圆盘费时多少？从n减去1费时多少？检验n是否等于1又费时多少？执行一次过程调用需要多少时间？因为我们只需要计算出运行时间的阶数，所以不必对此类问题作出精确的回答。但是我们必须将那些不依赖于n而只需常量时间的操作和那些运行时间依赖于n的操作区分开来。

一种可能是假设每种操作都取不依赖于n的常量时间。AHU（[1]）称这种假设为统一费用准则。按这种统一费用设想，令$T(n)$是n个圆盘的运行时间，我们得到差分方程

$$T(n)=2T(n-1)+C,$$

因为二次调用了对n－1个圆盘的相同的过程，且$C$是各种操作常量运行时间之和。设$T(1)$是对一个圆盘执行算法所需要的运行时间，用直接代入，可以得到

$$T(n)=(T(1)+C)2^{n-1}-C.$$

因为

$$\frac{T(1)}{2}2^n \leqslant T(n) < \left(\frac{T(1)+C}{2}\right)2^n,$$

因而

$$T(n)=\theta(2^n).$$

另一种可能是假设某些操作的运行时间是 n 的函数。但是，我们应当用一个 n 的什么函数呢？算法中出现的每一个数字都在 1 和 n 之间，圆盘也可以用 1 到 n 的数字来表示。而这样的数字可以用大约 log n 个毕特表之，因而有理由认为处理数字或圆盘的每次操作其运行时间是 log n 的一个常数倍。AHU([ 1 ])将此称为对数费用准则，并建议在一个算法所使用的数字没有明确的界限时使用它。按对数费用准则，我们得到算法运行时间的差分方程

$$T(n)=2T(n-1)+C\log n.$$

通过代入可以验证这个差分方程的解是

$$T(n)=2^n\left[\frac{T(1)}{2}+C\sum_{i=1}^{n}\frac{\log i}{2^i}\right].$$

用比较判别法可以证明在这个解中的累加和是收敛的，且可以假设它收敛于一个正的常数，我们有

$$T(n)=\theta(2^n).$$

因为两种费用准则都给出了相同的运行时间，我们得到

**命题2** 算法 HANOI 的运行时间为 $\theta(2^n)$。

尽管我们已经确定了解 Hanoi 塔问题的一个特定算法的运行时间，但是我们尚未确定问题的时间复杂度、我们需要确定一个下界，使得解此问题的每一个算法的运行时间都必须大于或等于这个下界。通过下面命题 3 的证明，我们确认 $\theta(2^n)$ 为问题的下界。

命题3 Hanoi 塔问题的时间复杂度为 $\theta(2^n)$。

证明 由命题1的证明可以直接导出解 Hanoi 塔问题所需要的最少移动次数为$2^n-1$。因为每次移动至少需要常量时间，因而我们已经确定了时间复杂度的下界。

该问题的时间复杂度的上界来自命题2。因为上、下界的阶数相等，因此我们把 $\theta(2^n)$ 定为该问题的时间复杂度。证毕。

现在我们已经知道了算法 HANOI 的时间复杂度，我们将转而考虑其空间复杂度。首先，我们将根据时间下界确定空间的一个下界。

命题4 解 Hanoi 塔问题的任何算法都至少需用（n+常数）个毕特的存储量。

证明 因为要解此问题，算法必须要产生 $2^n-1$次移动，必须要区分$2^n$ 种不同的状态。如果算法不区分这些状态，那么在两个不加区分状态的每一个之后，算法将经过相同的移动次数而停止，这至少在一种情况下产生错误。

一个算法所要区分的状态数等于在算法中存储状态的数目乘以内部状态数。因为算法有一个固定的有限规模，它只能有常数个不同的内部状态，而存储状态数是2的存储毕特数次幂。因而 $C\cdot 2^{\text{毕特数}}\geqslant 2^n$，于是毕特数$\geqslant n-\log C=n+$常量。证毕。

为了讨论递归算法的空间复杂度，现在让我们来考虑一下所使用的数据结构。两种可能使用的数据结构是数组和堆栈。数组是以一串连续整数为索引的地址集合，使得在数组中一个地址上所存储的信息可以通过指明此信息存储位置的索引整数来加以引用。例如在数组ARRAY中，位置I上的信

息可以通过ARRAY[I]来引用。堆栈是一个线性次序的位置集合，在堆栈中，信息的插入或删除只能在其顶端进行。

每个塔都可以用一个具有n个地址的数组来表示，且每个地址最多需要log n个毕特。因此一个数组数据结构有$\theta$(nlog n)个毕特就足够了。另外，每个塔也可以用一个堆栈来表示。栈中的每个地址将需要log n个毕特。因此，这也是一个有$\theta$(nlog n)毕特数的数据结构。事实上，我们还可以节省一点。因为只需要表示n个圆盘，因而堆栈结构仅需要n个地址，相比之下，数组结构就需要3n个地址。另一种可能的数据结构是一个数组，其第i个元素记录第i个圆盘所在的塔名。这种结构仅仅需要$\theta$(n)毕特数。还存在另一种可能性，即不去标记塔名，仅按从__到__的形式输出所作的移动，那我们对塔就可以不用费任何存储空间了。

递归算法仍然需要空间来当作其递归栈。当一个递归算法调用本身时，新的一轮调用的参数将取代上一轮参数的位置，而上一轮所用参数就需要先放到一个栈上，当新一轮调用完成后，这些参数可以再从栈被再次调用。返回地址当然也要放入堆栈中。对一次单独调用来说，称所有这些信息——参数和返回地址——为栈的一层结构。在任何一个时刻最多有n层结构是有效的，且每层结构用常数个毕特标记塔名，用log n个毕特标记圆盘数。因此不管塔名是否实际上被显示，一个递归算法将用$\theta$(nlog n)毕特。我们将上述内容归结为下面的命题。

**命题5**　递归算法HANOI正确地解了Hanoi塔问题，并用了$\theta(2^n)$时间和$\theta$(nlog n)空间。

递归算法所占用的空间大于最小空间。我们面临几种可

能性：

（1）最小空间仅仅是一个不能为任何算法所达到的下界。

（2）最小空间只能用一个超过最小时间的算法所达到。

（3）某些算法能同时达到最小时间和最小空间。通过改进一系列的迭代算法，我们将得到一个使用最小时间和最小空间的算法。

## 四、某些迭代算法

为了得到一个较好的算法，首先我们将考虑这样一个迭代算法：它模拟了n≥2时的递归算法。该算法与[7]中给出的一个算法类似，但是我们采用了一种明确追踪栈计数器的办法，这将有助于我们找一个使用更少空间的算法。

```
PROCEDURE RECURSIVE SIM(A,B,C,n)
  I:=1
  L1[1]:=A; L2[1]:=C; L3[1]:=B
  NUM[1]:=n-1; PAR[1]:=1; PAR[0]:=1
  WHILE I ≥1 DO
    IF NUM[I]>1
      THEN L1[I+1]:=L1[I]
           L2[I+1]:=L3[I]
           L3[I+1]:=L2[I]
           NUM[I+1]:=NUM[I]-1
           PAR[I+1]:=1
           I:=I+1
```

```
ELSE 从L1[I]到L3[I]移动
      WHILE PAR[I]=2 DO
            I:=I-1
      IF I≥1 THEN从L1[I]到L2[I]移动
                  PAR[I]:=2
                  TEMP:=L1[I]
                  L1[I]:=L3[I]
                  L3[I]:=L2[I]
                  L2[I]:=TEMP
```

塔名存储在三个数组 L1,L2,L3 中；在一次递归调用中的圆盘数存储在数组 NUM 中；PAR 值标明这是一对递归调用的第一次还是第二次调用。

RECURSIVE SIM 准备了调用 HANOI(A,C,B,n-1) 的各项参数。当该次调用的最后一个移动完成时，这些数组将存放了从 1 到n-2个圆盘调用的各项参数，每次这种调用都有 PAR=2。这些数组中仍然存放了 PAR=1 时 (A,C,B,n-1) 调用的各项参数。内层循环WHILE使 PAR=2的每次调用退出，保持数组计数器指针指向 (A,C,B,n-1) 调用。此时，因为 I=1，IF条件满足，则 "从L1[I]到 L2[I] 的移动" 完成了HANOI递归算法中的 "从A到C 的移动"。紧接着的赋值语句形成了PAR=2的 (B,A,C,n-1) 调用。因此，当这次调用中的移动完成时，数组中的所有调用都有PAR=2，且内层循环 WHILE 将置 I 为零，使所有这些调用退出。此时，IF条件不满足，故不执行任何操作，并且外层WHILE条件也不满足，因此算法终止。

命题6 RECURSIVE SIM算法在 $\theta(2^n)$时间和 $\theta(n\log n)$

空间内正确地解了Hanoi塔问题。

**证明** 因为该算法是已经证明了其正确性的递归算法的模拟，因此其正确性显而易见。主要是数组占用空间。因为每次I增值，而相应的NUM[I]减值，但NUM[I]绝不会小于1，因此在任何时候数组中最多有n－1个位置被占用。四个数组L1,L2,L3和PAR的每个元素仅占用常数量的空间，但NUM必须能存储n－1这么大的数，因此它的每个元素要占用$\theta(\log n)$毕特。于是，这些数组共占用$\theta(n\log n)$毕特。

现在我们来论证时间的占用。大多数操作涉及的是常数量大小的计算数，因此这些操作将取常数时间。只有增值，减值，赋值和比较数字的操作除外，它们所涉及的计算数可以有$\theta(\log n)$毕特。时间的差分方程是

$$T(n)=2T(n-1)+C\log n,$$

其中$T(n)$是解一个有n个圆盘的Hanoi塔问题所需时间，$C\log n$是操作具有$\theta(\log n)$毕特的计算数所费时间。与命题1的证明类似，我们有$T(n)=\theta(2^n)$。证毕。

请注意这个算法并没有对递归算法作任何改进，但通过对这种形式的研究，导致我们得到一个节省空间的办法。存储数组NUM引起了$\theta(n\log n)$空间的占用。如果我们不需要存储NUM，算法可以仅需$\theta(n)$空间。我们需要记录NUM吗？NUM被用作控制变量，因此看起来它是必要的。但是，如果注意到NUM[1]＋1，我们得到n，当NUM[I＋1]确定后，它一定等于NUM[I]－1，但此时

$$\begin{aligned}NUM[I+1]+I+1&=NUM[I]-1+I+1\\&=NUM[I]+I=n.\end{aligned}$$

因而我们所需要的有关NUM的信息被存储在I和n中。那么，

如果我们用测试 I = n - 1 来代替测试NUM[I] = 1，我们就可以避免存储NUM，并将空间复杂度从 $\theta$(nlog n)改进到$\theta$(n)。这种改变没有增加算法中任何一步的时间复杂度，因此其时间复杂度仍保持为 $\theta(2^n)$。

新的程序是

```
PROCEDURE NEW SIM (A,B,C,n)
  I:=1
  L1[1]:=A; L2[1]:=C; L3[1]:=B
  PAR[1]:=1; PAR[0]:=1
  WHILE I >1 DO
    IF I≠n-1
      THEN L1[I+1]:=L1[I]
           L2[I+1]:=L3[I]
           L3[I+1]:=L2[I]
           PAR[I+1]:=1
           I:=I+1
      ELSE从L1[I]到L3[I]移动
           WHILE PAR[I]=2 DO
                 I:=I-1
           IF I ≥ 1 THEN从L1[I]到L2[I]移动
                         PAR[I]:=2
                         TEMP:=L1[I]
                         L1[I]:=L3[I]
                         L3[I]:=L2[I]
                         L2[I]:=TEMP
```

从上面的观察，我们有

**命题7** NEW SIM算法在$\theta(2^n)$时间和$\theta(n)$空间内正确地解了Hanoi塔问题。

尽管已经达到$\theta(n)$的空间复杂度，我们还希望进一步降低它而达到（n+常数）毕特。我们发现算法扫描PAR找到第一个不等于2的元素，然后将此元素置为2并将此元素前的所有为2的元素置为1。这与我们熟悉的加1到一个二进制数上的操作极为相似：发现的第一个零位，我们用1来代之，并将其前面所有的1位换为零。因此，看起来我们能够用一个简单的计数器来代替数组PAR。当然计数器中的毕特数依赖于n。

至此，尚未导致任何空间的节省。能从计数器中得到足够的信息来确定我们将从哪个塔上移动圆盘吗？一个肯定的回答将使我们得到一个最小空间的算法。为了启发我们设计这个最小空间的算法，我们将检查解5个圆盘的Hanoi塔问题所需要的31次移动序列。该序列在表1中列出。

在问题的解中，每隔一次都要移动盘1，所以如果我们知道了哪个塔上有盘1，就知道了从哪个塔作移动，但还不知道要移动到哪个塔上去。我们考虑将三个塔排成一个圆圈，从表1可以看出，奇数个圆盘的Hanoi塔问题，盘1总是沿逆时针方向移动。类似地，当圆盘数为偶数时，盘1总是按顺时针方向移动。于是，追踪含圆盘1的塔，那么不管n是奇数还是偶数，我们都能知道怎样做其他的移动。

对于那些不涉及圆盘1的移动，我们知道移动只在不含盘1的那两个塔上进行。再看表1，我们看到标号为奇数的盘总与盘1的移动方向一致，而标号为偶数的盘总与之相反。所以知道了所涉及的塔名和将要移动的圆盘的标号奇偶

表1　五个圆盘的Hanoi塔问题的解

| 塔 0 | 塔 1 | 塔 2 | 十进制计数器 | 计数器 | 圆盘 | 从 | 到 |
|---|---|---|---|---|---|---|---|
| 12345 | - | - | 0 | 00000 | 1 | 0 | 2 |
| 2345 | - | 1 | 1 | 00001 | 2 | 0 | 1 |
| 345 | 2 | 1 | 2 | 00010 | 1 | 2 | 1 |
| 345 | 12 | - | 3 | 00011 | 3 | 0 | 2 |
| 45 | 12 | 3 | 4 | 00100 | 1 | 1 | 0 |
| 145 | 2 | 3 | 5 | 00101 | 2 | 1 | 2 |
| 145 | - | 23 | 6 | 00110 | 1 | 0 | 2 |
| 45 | - | 123 | 7 | 00111 | 4 | 0 | 1 |
| 5 | 4 | 123 | 8 | 01000 | 1 | 2 | 1 |
| 5 | 14 | 23 | 9 | 01001 | 2 | 2 | 0 |
| 25 | 14 | 3 | 10 | 01010 | 1 | 1 | 0 |
| 125 | 4 | 3 | 11 | 01011 | 3 | 2 | 1 |
| 125 | 34 | - | 12 | 01100 | 1 | 0 | 2 |
| 25 | 34 | 1 | 13 | 01101 | 2 | 0 | 1 |
| 5 | 234 | 1 | 14 | 01110 | 1 | 2 | 1 |
| 5 | 1234 | - | 15 | 01111 | 5 | 0 | 2 |
| - | 1234 | 5 | 16 | 10000 | 1 | 1 | 0 |
| 1 | 234 | 5 | 17 | 10001 | 2 | 1 | 2 |
| 1 | 34 | 25 | 18 | 10010 | 1 | 0 | 2 |
| - | 34 | 125 | 19 | 10011 | 3 | 1 | 0 |
| 3 | 4 | 125 | 20 | 10100 | 1 | 2 | 1 |
| 3 | 14 | 25 | 21 | 10101 | 2 | 2 | 0 |
| 23 | 14 | 5 | 22 | 10110 | 1 | 1 | 0 |
| 123 | 4 | 5 | 23 | 10111 | 4 | 1 | 2 |
| 123 | - | 45 | 24 | 11000 | 1 | 0 | 2 |
| 23 | - | 145 | 25 | 11001 | 2 | 0 | 1 |
| 3 | 2 | 145 | 26 | 11010 | 1 | 2 | 1 |
| 3 | 12 | 45 | 27 | 11011 | 3 | 0 | 2 |
| - | 12 | 345 | 28 | 11100 | 1 | 1 | 0 |
| 1 | 2 | 345 | 29 | 11101 | 2 | 1 | 2 |
| 1 | - | 2345 | 30 | 11110 | 1 | 0 | 2 |
| - | - | 12345 | 31 | 11111 | | | |

性，我们就可以决定要做怎样的一次移动。

从计数器中我们能确定将要移动的圆盘号是奇的还是偶的吗？让我们看看表1中“计数器”列的内容：其最右零的位置告诉了我们将要移动的盘号。于是，一个有n个毕特的

计数器对于解Hanoi塔问题已经足够了。

我们用以上事实构造下述算法。

```
PROCEDURE TOWERS(n)
  T:=0（*塔号按模3(mod 3)计算*）
  COUNT:=0（*COUNT(计数器)有n个毕特*）
       ⎧  1   如果n是偶数
  P:=  ⎨
       ⎩ -1   如果n是奇数
  WHILE TRUE DO
    从T到T+P移动圆盘1
    T:=T+P
    COUNT:=COUNT+1
    IF COUNT=ALL 1's THEN RETURN
    IF COUNT中最右零是在偶的位置
      THEN从T-P到T+P移动圆盘
      ELSE从T+P到T-P移动圆盘
    COUNT:=COUNT+1
  ENDWHILE
```

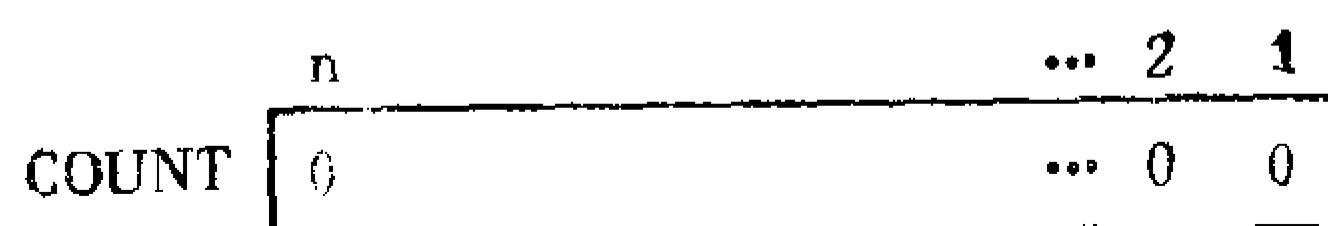



图2 计数器（COUNT）所用的存储示意图。注意：COUNT包含n个毕特，将最右位的毕特称为位置1，从右到左的位置依次为奇、偶、奇、偶…

**附注** 我们还可以通过去掉第一个COUNT:=COUNT+1语句并删除COUNT的最右位毕特来进一步改进这个算法。

这将需要改变 COUNT 中毕特的编号，使得它最右位毕特是毕特零。T.R.崴尔施（[8]）最近发表了一个与TOWERS程序相似的算法。

我们必须证明TOWERS算法正确地解了 Hanoi 塔问题。证明是这样实现的：当 COUNT 存放了一个形如 $2^k-1$ 的数时，我们证明一串确定的移动序列已经完成，使得当 $k=n$ 时，对HANOI(A,B,C,n) 的移动序列完成并且程序终止，因为此时COUNT中每一位均为 1。

**命题8** 当 $COUNT=2^k-1$ 时，即COUNT = | 00…01…1 | 中有k个 1，那么

如果 $k\not\equiv n \pmod 2$，HANOI(A,C,B,k) 的正确移动已经完成，且T的内容是1（代表B）；

如果 $k\equiv n \pmod 2$，HANOI(A,B,C,k) 的正确移动已经完成，且T的内容是2（代表C）。

**证明** 若 $k=1$，只完成一次T到T+P的移动，也就是当n为奇数时，这是A到C的移动，当n为偶数时，则为A到B；且T的内容是T+P，即当n为奇数时，T+P是2，当n为偶数是T+P为1。这正符合我们的论断。

注意，紧接在 IF…RETURN 语句之前，COUNT 只能取值 $2^k-1$。假设对HANOI(A,B,C,k) 或 HANOI(A,C,B,k) 的移动已经完成，若 $k\not\equiv n \pmod 2$，根据假设现在 T 的内容是 1，因此将是A到 C 的移动：如果 k 是奇数，该次移动是由T－P到T＋P，即从1－(1) 到1＋1，表示由A到C；若k是偶数，移动是由T＋P到T－P，即从1＋(－1) 到1－(－1)，表示由A到C。若 $k\equiv n \pmod 2$，下一次移动将是A到B，因为根据假设现在T的内容是2；若k为奇的，该次移动为T－P到

T + P，即2 - ( - 1) 到2 + ( - 1)，表示A到B的移动；若 k为偶的，那么移动为T + P到T - P，即2 + 1 到 2 - 1，表示由 A 到B的移动。

紧接在下面的 COUNT 将增至 $\boxed{0\cdots010\cdots0}$，即尾部含k 个零。当 COUNT = $2^{k+1}$ - 1 时，算法将重复如前的一串移动，这是因为它只“看” COUNT 中最右位所含信息，存在的差别在于 T 在开始时有不同的值。T 的不同的起始值将引起标号的循环排列。如果 $k \not\equiv n \pmod 2$，那么所完成的移动将是

HANOI(A,C,B,k)

A 到 C

HANOI(B,A,C,k),

这导致了 $k+1 \equiv n \pmod 2$ 的 HANOI (A,B,C,k + 1)，并且T 的内容将是2（即 1 + 1）．如果 $k \equiv n \pmod 2$，那么完成的移动将是

HANOI(A,B,C,k)

A 到 B

HANOI(C,A,B,k),

这导致了 $k+1 \not\equiv n \pmod 2$ 的 HANOI(A,C,B,k + 1)，并且T 的内容将是 1(即 2 + 2)。证毕。

**命题 9** TOWERS 算法占用了 $\theta(2^n)$时间和(n + 常量)毕特空间。

**证明** 对于占用空间，在 COUNT 中有 n 个毕特，而 T 和 P 只占用常量毕特。

对于时间，初始部分花费 $\theta(n)$且 WHILE 循环被迭代了 $2^n - 1$ 次，若每次迭代花费常量时间，我们将有 $\theta(2^n)$。

但是在计数器上所作的测试和增值指令要花费$\theta(n)$时间，这导致时间为$\theta(n2^n)$。因此我们必须要进一步证明它仅占用$\theta(2^n)$时间。

如果COUNT中的值是偶的，那么增值和测试仅只需要观察一位毕特。若COUNT中的值是奇的，且(COUNT－1)/2是偶的，那么算法仅需观察二个毕特。事实上，算法在$2^{n-k}$种情况中将观察COUNT的k位毕特。因为$\sum_{k=1}^{\infty} k \cdot 2^{-k}$收敛，所以占用的时间将是$\theta\left(\sum_{k=1}^{n} k \cdot 2^{n-k}\right)=\theta(2^n)$。证毕。

我们将这些结果概括为下述定理。

**定理** 解有n个圆盘的Hanoi塔问题的任何一个算法至少需占用$\theta(2^n)$时间和（n＋常量）毕特的存储量。算法TOWERS解了这个问题并且同时达到了最少时间和空间。

## 五、概述和总结

得到最好算法的目的已经达到了，为了达到这个目标，我们先以“分而治之”的方式入手分析这个问题，并且从此分析中得出一个能够证明其正确性的递归算法。其次，我们分析了这个递归算法所用的时间。因为这个问题的任何一个解都需要$2^n-1$次移动，因此这个时间是最佳可行的需求。

由时间的下界，我们导出解此问题的任何算法所需空间的下界为n个毕特。对递归算法的空间分析指出其所用空间大于我们的下界，且递归栈需要占用这样大的空间。

为了降低对空间的需求，我们构造了一个直接模拟递归算法的迭代算法。因为这是一个直接的模拟，它占用了相同

数量的空间，但是我们可以研究被存储的所有信息是否是必要的，并发现了存储每次递归调用的圆盘数是不必要的。这导致产生了一个新的只占用$\theta(n)$空间的迭代算法。

然后，我们注意到有一个数组正充当着一个计数器的作用，但是用一个计数器来代替它并不能减少空间占用量。紧接着我们又研究了计数器中是否有足够的信息告诉我们移动哪个盘和移到哪儿去。我们发现它可以告诉我们要移动哪个盘，但移动到哪个塔上这取决于圆盘总数是奇的还是偶的。

当增加两个变量分别用来记录圆盘数的奇偶性及含有最小圆盘的塔名时，我们发现已经有了足够的信息来解决这个问题了，而且还可以去掉那些在每一次模拟的递归调用中记录塔名的数组。

现在，我们已经有了一个占用空间为（n＋常数）毕特的算法，这个量等于空间的下界。然后，我们证明了这个算法仍然具有最少时间阶，所以得到了一个最好的算法。

需要提请注意的是，还可能存在其他的一些看起来很不相同的算法，这些算法均在最少时间和空间内解决了此问题。本文试图通过示范给出一种人们在推导一个好的算法时常用的设计方法。我们之所以选择 Hanoi 塔问题是因为对此问题可以导出最好的算法，而对别的问题这可能是很难的。很可能我们会找到一个可证明其正确性的算法及对此问题的下界，但却发现在算法的运行时间和下界之间存在着差距，或发现算法的占用空间量和其下界之间存在差距。经常出现这样的情况，对问题最关键一步的认识，如 Hanoi 塔问题中对圆盘按顺时针或逆时针移动的认识，在一个算法给出后的许多年间都可能未被发现。另一方面，一个已给出的

算法可能是一个最好的算法，但为提高该问题的下界仍需有一个深刻的认识。

无论如何，我们希望已经给读者一些关于算法分析的感性认识。

## 六、练　　习

为了检验自己是否真正掌握了这一技巧，对类似的问题试着用一用这个技巧，将是有益的。在此，我们给出另外两个解 Hanoi 塔问题的算法。如果您决定试一试的话，您的任务是证明这些算法确实解决了这个问题(即证明其正确性)，并确定这些算法所占用的时间和空间。

练习 1：

```
PROCEDURE HANOI ITERATIVE (A,B,C,n)
  IF n mod 2=0 THEN MOVE[1]:=A TO B
               ELSE MOVE[1]:=A TO C
  K:=1
  WHILE n>1 DO
    n:=n-1;K:=2*K
    IF n mod 2  =0 THEN MOVE[K]:=A TO B
                        L1:=C;L2:=A;L3:=B
                   ELSE MOVE[K]:=A TO C
                        L1:=B;L2:=C;L3:=A
    FOR I:=1 TO K-1 DO
        CASE MOVE[I] OF
              A TO B:MOVE[K+I]:=L1 TO L2
              A TO C:MOVE[K+I]:=L1 TO L3
```

B TO A:MOVE[K + I]:=L2 TO L1

B TO C:MOVE[K + I]:=L2 TO L3

C TO A:MOVE[K + I]:=L3 TO L1

C TO B:MOVE[K + I]:=L3 TO L2

**提示**：为证明其正确性，您不妨这样想：引进一个新变量，并证明在WHILE循环的每一次迭代中，这个新变量增值（或者，您想减值也行），在每次迭代终结时，一个规模依赖于此新变量的Hanoi塔问题已被解决。您将需要给出已被解决问题的那些塔名，还需要给出新变量的值。

对于空间，您应当知道，算法将每一次移动存储在数组MOVE中。

对于时间，您不妨兼而考虑统一的及对数的费用准则。

**练习2**（见[3]）：

按顺时针方向将最小圆盘移动到一个塔上。

<u>WHITE</u> 一个圆盘（不是最小的）能够被移动 <u>DO</u>

　　移动这个圆盘

　　按顺时针方向将最小圆盘移动到一个塔上

<u>ENDWHITE</u>

**提示**：对正确性证明您须小心行事，因为此算法只给出了当圆盘数是偶数时的原始Hanoi塔问题的解。您可能会想到引进一个新变量，并证明当完成一定次数的移动后，圆盘的排列状态是一个您给出的新变量的特定函数。

对于时间和空间，上述算法是不完全的，因为它没有指定一种数据结构用以确定一个圆盘是否能被移动。您可以考虑对每一个塔用一个整数栈表示，栈中整数代表在这个塔上的圆盘。另外，您还可以考虑用一个数组DISK来表示这些

信息，使得 DISK[I]的内容是第 I 个 最大圆盘所在的塔名。您也可能发现这对于证明第 i 个圆盘被移动了 $2^{n-i}$ 次是很有用的。

## 参 考 文 献


[1] A. Aho, J. Hopcroft, and J. Ullman, The Design and Analysis of Computer Algorithms, Addison-Wesley, Reading, MA, 1974.

[2] S. Baase, Computer Algorithms, Addison-Wesley, Reading, MA, 1978.

[3] P. Buneman and L. Levy, The Towers of Hanoi Problem. Information Processing Letters 10, 1980, p. 243—244.

[4] E. Horowitz and S. Sahni, Fundamentals of Computer Algorithms, Computer Science Press, Rockville, MD, 1978.

[5] D. Knuth, The Art of Computer Programming, vol. 1, Fundamental Algorithms (2nd ed.1973) , vol.2, Seminumerical Algorithms (2nd ed., 1981) , vol. 3, Search and Sorting (1973), Addison-Wesley, Reading. MA.

[6] H. Simon, The Functional Equivalence of Problem Solving Skills, Cognitive Psychology, 1975, p. 268—288.

[7] A. Tenenbaum and M. Augenstein, Data Structures Using PASCAL, Prentice-Hall, Englewood Cliffs, NJ, 1981, p. 149—154.

[8] T.R. Walsh, The Towers of Hanoi Revisited:Moving the Rings by Counting the Moves, Information Processing Letters 15, 1982, p.64—67.

[9] D. Wood,The Towers of Brahma and Hanoi Revisited, Computer Science Technical Report No. 80-CS-23, McMaster University, 1980.


（刁在筠编译，方祖耀校）

# 修改了的迭代及概率①

Lawrence J. Wallen

数学上经常会出现这样的情况：某个数学分支中的定理，实际上证明，推广、简化，或者“解释”了另一截然不同的分支中的某个定理。这种数学上的交叉当然是多多益善的。概率论与经典分析之间的这种交叉也是数不胜数的。本文又给出了这种交叉之一例。

在适当条件下，函数的迭代会产生不动点。但一般的迭代过程则可能具有这样的不良性质：要么收敛于不动点的速度较慢，要么根本不收敛于不动点。本文正是要对具此不良性的迭代过程加以修正，并用纯分析的方法找出不收敛的原因。我们的修正实际上也是机会博奕迭代中的非最优化方法。由本文中所给出的关于点点收敛的定理，不难推知分析中的应用准则。

我们先给出分析上的表述。设 $G(w)$ 是区间[0,1]上连续可微的严格凸函数，满足 $G(0)>0$，$G(1)=1$，且当 $0\leqslant w<1$ 时，$0\leqslant G'(w)<1$。记 $\eta_0=0$，$\eta_{n+1}=G(\eta_n)$，则知 $\eta_n\to 1$。即 $\eta_n$ 收敛于 $G$ 的不动点 1 。

设 $w_n$ 为满足 $0\leqslant w_n\leqslant 1$ 的序列。利用 $\eta_n$ 可得序列 $\xi_n$，

---

① Modified iteration and probability, *AMM*, 92(1985), 481—485.

$\xi_0=0$,

$$\xi_{n+1}=G(w_{n+1})+(\xi_n-w_{n+1})G'(w_{n+1}),$$

即 $(\xi_n,\xi_{n+1})$ 在 $G$ 的过点 $(w_{n+1},G(w_{n+1}))$ 的切线上。由 $G$ 之严格凸性知 $\xi_{n+1}\leqslant G(\xi_n)$，且等号成立的充要条件为 $w_{n+1}=\xi_n$。因 $G$ 为严格增函数，故 $\xi_n\leqslant\eta_n$。特别地，取 $w_{n+1}\equiv\xi_n\equiv\eta_n$，即得通常的迭代。

那么，序列 $\{w_n\}$ 满足什么条件才能保证 $\xi_n\to1$ 呢？图象上是显然的，即 $w_n$ 必须趋于 1。

我们对 $G'(1)=1$ 的情况感兴趣。但在此假设下，如果 $w_{n+1}=1$，则 $\xi_{n+1}=\xi_n$，从而推知 $\xi_n$ 不趋于 1。由此可见，只有当 $w_n$ 趋于 1 的速度不太快时才可能有 $\xi_n\to1$。这正是下列定理的内容。

**定理 1** $\xi_n\to1$ 之充分必要条件是，$w_n\to1$ 且

$$\sum_{n=0}^{\infty}(1-G'(w_n))=\infty.$$

**证明** 必要性 设 $1-\xi_n\to0$，要证上述级数发散，或等价地，要证 $\prod_{n=1}^{\infty}G'(w_n)=0$。由

$$\begin{aligned}1-\xi_{n+1}&=1-G(w_{n+1})-(\xi_n-1+1-w_{n+1})G'(w_{n+1})\\&=1-[G(w_{n+1})+(1-w_{n+1})G'(w_{n+1})]\\&\quad+(1-\xi_n)G'(w_{n+1})\end{aligned}$$

及 $G$ 之严格凸性知，括号中项的最大值不超过 1。因此有

$$1-\xi_{n+1}\geqslant(1-\xi_n)G'(w_{n+1}),$$

从而

$$(1-\xi_{n+1})\geqslant(1-\xi_0)\prod_{j=1}^{n+1}G'(w_j),$$

由此推知$\prod_{n=1}^{\infty}G'(w_n)=0$。

**充分性** 由$G(w_n)>1-\varepsilon$知

$$G(w)-wG'(w)\geqslant w(1-G'(w))。$$

取$N$充分大，使当$n\geqslant N$时$w_n>1-\varepsilon$。于是当$n\geqslant N$时，

$$\begin{aligned}\xi_n-\xi_{n-1}G'(w_n)&=G(w_n)-w_nG'(w_n)\\&\geqslant(1-\varepsilon)(1-G'(w_n)),\end{aligned}$$

$$\xi_{n-1}-\xi_{n-2}G'(w_{n-1})\geqslant(1-\varepsilon)(1-G'(w_{n-1})),$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\xi_N-\xi_{N-1}G'(w_N)\geqslant(1-\varepsilon)(1-G'(w_N))。$$

在第二个不等式两边同乘以$G'(w_n)$，第三个不等式两边同乘以$G'(w_n)G'(w_{n-1})$，…，最后一个不等式两边同乘以$G'(w_n)G'(w_{n-1})\cdots G'(w_{N+1})$，分别相加，两边各项抵消后得

$$\xi_n-\xi_{N-1}\prod_{j=N}^{n}G'(w_j)\geqslant(1-\varepsilon)\left[1-\prod_{j=N}^{n}G'(w_j)\right]。$$

令$n\to\infty$，由$\prod_{N}^{\infty}G'(w_j)=0$知$\lim\inf\xi_n\geqslant1-\varepsilon$。

由此定理立得下列推论：

**推论1** (a) $\sum(1-G'(\eta_n))=\infty$，

(b) 若当$0\leqslant w<1$时$G''(w)\leqslant M$且$G'(1)=1$，则

$$\sum(1-\eta_n)=\infty。$$

**证明** 令$w_{n+1}=\eta_n$，由$\xi_n=\eta_1\to1$得(a)；对于(b)，由中值定理知$1-G'(\eta_n)\leqslant M(1-\eta_n)$，再依(a)即得。

现在我们把这一问题与概率论联系起来。迭代博奕进行如下：预先给定数列 $w_1, w_2, \cdots, 0 \leqslant w_n \leqslant 1$。从[0,1]区间中随机抽取 $X_1$，$X_1$ 具有连续分布函数 $F(x)$。赢值 $Y_1 = X_1$，若 $X_1 > w_1$；$Y_1 = 0$，若 $X_1 \leqslant w_1$。抽取 $X_2$（其分布函数仍为 $F(x)$），赢值 $Y_2 = X_2$，若 $X_2 > w_2$；否则 $Y_2 = Y_1$。依此重复下去。

严格地说，即$\{X_n, n \geqslant 1\}$是相互独立同分布的随机变量列，共同分布函数为满足 $F(0) = 0$，$F(1) = 1$ 且在[0,1]区间上严格增加的连续函数；而随机变量列$\{Y_n, n \geqslant 1\}$的定义为 $Y_0 = 0$，

$$Y_{n+1} = \begin{cases} X_{n+1}, & 若 X_{n+1} > w_{n+1}, \\ Y_n, & 若 X_{n+1} \leqslant w_{n+1}. \end{cases}$$

称 $w_n, n \geqslant 1$ 为门限数列，它们是在博奕开始前取定的。

那么，如何取 $w_1, w_2, \cdots, w_n$ 才能使 $E(Y_n)$ 最大呢（$E$ 为数学期望）？这一问题很早就被提出来了。计算 $E(Y_{n+1})$ 得：

$$E(Y_{n+1} | X_{n+1} = \xi) = \begin{cases} \xi, & 当 \xi > w_{n+1}, \\ E(Y_n), & 当 \xi \leqslant w_{n+1}. \end{cases}$$

因此，

$$\begin{aligned} E(Y_{n+1}) &= \int_0^{w_{n+1}} E(Y_{n+1} | X_{n+1} = \xi) \mathrm{d}F(\xi) \\ &\quad + \int_{w_{n+1}}^1 E(Y_{n+1} | X_{n+1} = \xi) \mathrm{d}F(\xi) \\ &= E(Y_n) E(w_{n+1}) + \int_{w_{n+1}}^1 \xi \mathrm{d}F(\xi). \end{aligned}$$

令 $E(Y_n) = \xi_n$ 及 $G(w) = 1 - \int_w^1 F(\xi) \mathrm{d}\xi$，对上式分部积分

得

$$\xi_{n+1}=G(w_{n+1})+(\xi_n-w_{n+1})G'(w_{n+1}),$$

这是因为 $G'=F$。因 $F$ 严格增加，故 $G$ 正是我们前边所考虑的函数。由于 $\xi_n\leqslant\eta_n$，故 $E(Y_n)$ 之最大值 $\eta_n$ 在取 $w_{j+1}=\eta_j=E(Y_j)$，$j=0,\cdots,n-1$ 时达到。可见，若博奕无限次地重复下去，则赢值 $(Y_n)$ 趋于 1。

进一步要问，若由于某种原因(比如，关于 $w_n$，$n\geqslant1$ 的选取可能有附加条件)而没有采取最优对策，如何使赢值趋于 1 呢?

下列定理彻底回答了这一问题，并用概率方法给出了定理 1 的证明。

**定理2** 设 $w_n\to1$，则下列事实等价：

(a) $Y_n\to1$ 以概率 1 成立(a.s)；

(b) $E(Y_n)\to1$；

(c) $\sum(1-G'(w_j))=\sum(1-F(w_j))=\infty$。

**证明** 设 $(\Omega,p)$ 为此过程的概率空间。对 $\omega\in\Omega$，令

$$N'(\omega)=\{v:X_v(\omega)>w_v\},$$

由 Borel-Cantelli 引理[1,p.41]知，集合 $N'$ 以概率 1 为有穷集或无穷集，若

$$\sum P(X_n>w_n)=\sum(1-F(w_n))$$

发散或收敛。

对每个 $n$，令 $v(n)$ 为集合 $N'$ 中满足 $v(n)\leqslant n$ 的最大数(当 $N'(\omega)=\varnothing$，令 $X_0=0$，$v(n)\equiv0$)，从而 $Y_n=X_{v(n)}$。于是，当

$$\sum(1-F(w_j))=\infty,$$

由 $w_n \to 1$ 知 $Y_n$ 几乎处处收敛到 1。

又令 $M=\max\{v: v\in N'(\omega)\}$，于是当

$$\sum(1-F(w_j))<\infty,$$

有 $\lim Y_n(\omega)=X_M(\omega)$，依有界收敛定理推知

$$\lim E(Y_n)=E(X_M)<1.$$

注意，在任何情形下，序列 $\xi_n=E(Y_n)$ 都收敛。

事实上，在以上证明中，我们也得到了在 $G'(0)=0$，$G'(1)=1$ 情形下的定理 1，不难得到在其它情形下关于定理 1 的证明。

现在假设我们可以依后验知识来选取 $w_n$（和相应的 $Y_n$），即取

$$w_n=w_n(X_1,X_2,\cdots,X_{n-1})$$

而 $Y_n$ 的定义同前。这时，最佳对策是显然的：

$$\hat{w}_n=\max(X_1,X_2,\cdots,X_{n-1}).$$

令 $\hat{Y}_n$ 为相应的 $Y_n$（以 $\hat{Y}_n=0$，$\hat{w}_1=0$ 为初值）。用简单的归纳方法可得，

$$\hat{Y}_n=\max(X_1,\cdots,X_n).$$

对任何其它选择 $(w_n,Y_n)$，因 $Y_n$ 是 $X_1(w),\cdots,X_n(w)$ 中的一个，于是有 $\hat{Y}_n\geqslant Y_n$，从而 $E(\hat{Y}_n)\geqslant E(Y_n)$。特别地，$E(\hat{Y}_n)\geqslant\eta_n$。

不难算得（见(1)式）

$$E(\hat{Y}_n)=1-\int_0^1 F^n(w)\,\mathrm{d}w,$$

从而

$$1-\eta_n \geqslant \int_0^1 [G'(w)]^n \mathrm{d}w.$$

也可用分析方法得此结果。记 $G_j$ 为$G$的 $j$ 重复合，注意到 $G(w) \geqslant w$，我们有

$$1-\eta_n = \int_0^1 \frac{\mathrm{d}}{\mathrm{d}w} G_n(w)\mathrm{d}w$$

$$= \int_0^1 \prod_{j=0}^{n-1} G'(G_j(w))\mathrm{d}w$$

$$\geqslant \int_0^1 [G'(w)]^n \mathrm{d}w,$$

这使我们更加清楚地看出推论 1 的 结 论(b)，即当

$$\int_0^1 \frac{\mathrm{d}w}{1-G'(w)} = \infty \text{时，} \sum(1-\eta_n) = \infty.$$

下面，我们把依最优预先选择得到的 $w_n = \eta_{n-1}$ 的期望值 $\eta_n$ 与利用后验知识选取的最优对策的期望值 $E(\hat{Y}_n)$加以比较，并以此结束本文。

固定 $n$，令 $X_n^1, X_n^2, \cdots, X_n^n$ 为 $X_1, \cdots, X_n$的顺序统计量，即

$$X_n^1 = \min(X_1, X_2, \cdots, X_n),$$

$X_n^j$ 为 $X_1, \cdots, X_n$依从小到大顺序排列后的第 $j$ 个值(见文献[3],ch.9)，$X_n^k$ 之期望由以下公式给出：

$$E(X_n^k) = n\binom{n-1}{k-1}\int_0^1 F^{-1}(u)u^{k-1}(1-u)^{n-k}\mathrm{d}u. \qquad (1)$$

考虑下列分布族

$$F_p(w) = 1 - (1-w)^p, \qquad p>0,$$

则相应于 $G_p(w)$ 的 $\eta_n$ 的渐近性质为

$$1-\eta_n \sim \left(\frac{1+1/p}{n}\right)^{1/p}.$$

文献[4]的第 223 页给出了此结论的简单证明。将 $F_p$ 代入(1)式，经常规计算知

$$E(X_n^k) = n\binom{n-1}{k-1}\int_0^1 (1-(1-u)^{\frac{1}{p}})u^{k-1}(1-u)^{n-k}du$$

$$= 1 - \frac{\Gamma(n+1)}{\Gamma\left(n+\frac{1}{p}+1\right)} \cdot \frac{\Gamma\left(n-k+\frac{1}{p}+1\right)}{\Gamma(n-k+1)}. \qquad (2)$$

设 $n-k$ 为固定值，如 $n-k+1=\theta$，由

$$\Gamma(n+z) \sim n^z \Gamma(n) \quad (见[2]，P.212),$$

我们希望找出满足 $1-E(X_n^k) \sim 1-\eta_n$ 的 $k$ 值，或利用(1)，(2)式解出

$$\frac{\Gamma\left(\theta+\frac{1}{p}\right)}{\Gamma(\theta)} = \left(\frac{p+1}{p}\right)^{\frac{1}{p}}.$$

下面是一些解(后三个解是近似解)：

$p=1$（均匀分布），　　$k=n-1$；

$p=\frac{1}{2}$，　　$k=n-1.50$；

$p=\frac{1}{3}$，　　$k=n-2.08$，

$p=2,\qquad\qquad k=n-0.73.$

在均匀分布下，使用最优预先选择可以得到更为优越的结果：与第二个大的$X$之期望一致，当$p=\frac{1}{3}$，与第三个大的$X$的期望一致，如此等等。这似乎意味着，当取较大值的概率更大时，用后验选择更为有效。

## 参考文献


[1] L.Breiman, Probability, Addison-Wesley, Reading, MA, 1968.

[2] E.T.Copson, An Introduction to the Theory of Functions of a Complex Variable, Clarendon Press, Oxford, 1935.

[3] S.Karlin, A First Course in Stochastic Processes, Academic Press, New York, 1966.

[4] G. Szekeres, Regular iteration of real and complex functions, *Acta Math.*, 100 (1958) 203—258.


（李东风译，范永亮、谢衷洁校）

# 美国第47届 Putnam 数学竞赛试题与解答①

L.E. Klosinski, G.L. Alexanderson,
L.C. Larson

Putnam 数学竞赛是由美国数学会主办的一个年度竞赛，参加者为美国及加拿大的大学生。它是由 William Lowell Putnam 基金赞助的，这笔基金是 Putnam 夫人为纪念她的丈夫而捐助的．以奖励竞赛中的获胜者。

第47届Putnam 竞赛于1986年 9 月 6 日举行，有来自美国和加拿大的 358 个院校的 2094 位大学生参加了竞赛。六名最高分的优胜者被吸收为 Putnam 会员。集体一等奖为哈佛大学数学系获得。

第47届 Putnam 竞赛的命题委员会有以下人员：Richard P.Stanley，Abraham P.Hillman 和 Harold M.Stark。本文介绍这次竞赛的问题和解答。

## 问　　题

**问题 A-1**　设 $f(x)=x^3-3x$，设 $S$ 是所有满足 $x^4+36\leqslant 13x^2$ 的实数的集合。求

① William Lowell Putnam mathematical competition, *Amer. Math. Monthly*, 94 (1987) 747—756.

$\max\limits_{x\in S} f(x)$，要求说明理由。

**问题 A-2** 用$[x]$表示不超过$x$的最大整数。求出$[10^{20000}/(10^{100}+3)]$的个位数字。

**问题 A-3** 求$\sum\limits_{n=0}^{\infty}\text{Arccot}(n^2+n+1)$的值，这里，当$t\geqslant 0$时，$\text{Arccot}\, t$ 确定 数 $\theta\in\left(0,\dfrac{\pi}{2}\right]$，满足 $\cot\theta=t$。

**问题 A-4** 称 $n\times n$ 矩阵$A$的 $n$ 个元素 为$A$的一个"横截"，这 $n$ 个元素中的任意二个都不在同一行，也不在同一列。

用$f(n)$表示满足下面两条件的 $n\times n$ 矩阵$A$的个数：

(a) $A$的任一元素 $a_{ij}\in$集合$\{-1,0,1\}$。

(b) 对$A$的所有的横截，其 $n$ 个元素之和恒相等。

例如，下面的$A$满足要求：

$$A=\begin{pmatrix}-1 & 0 & -1\\ 0 & 1 & 0\\ 0 & 1 & 0\end{pmatrix}.$$

请对$f(n)$确定一个形如

$$f(n)=a_1b_1^n+a_2b_2^n+a_3b_3^n+a_4$$

的公式，此处 $a_i$ 和 $b_i$ 均为有理数。请证明你的公式。

**问题 A-5** 设 $f_1(x),f_2(x),\cdots,f_n(x)$ 都是 $n$ 元函数，$x=(x_1,x_2,\cdots,x_n)\in R^n$（$R$ 为实数域），满足以下条件：

(1) 对 $1\leqslant i\leqslant n$，$f_i(x)$在 $R^n$ 上处处存在二阶连续偏导数。

(2) 对 $1\leqslant i\leqslant n$，$1\leqslant j\leqslant n$，存在常数 $c_{ij}$，使得

$$\frac{\partial f_i}{\partial x_j}-\frac{\partial f_j}{\partial x_i}=c_{ij}.$$

证明在 $R^n$ 上存在一个函数 $g(x)$，对 $1\leqslant i\leqslant n$，使得 $f_i+\partial g/\partial x_i$ 是线性函数(形如

$$a_0+a_1x_1+a_2x_2+\cdots+a_nx_n$$

的函数，称为线性函数)。

问题A-6　设 $a_1,a_2,\cdots,a_n$ 是实数，$b_1,b_2,\cdots,b_n$ 是不同的正整数。

设有多项式 $f(x)$ 满足下面的恒等式

$$(1-x)^nf(x)=1+\sum_{i=1}^{n}a_ix^{b_i}. \qquad (*)$$

对 $f(1)$ 求一个简单的表达式，满足：(1) 只依赖 $b_1$, $b_2,\cdots,b_n$ 和 $n$，与 $a_1,a_2,\cdots,a_n$ 无关。(2) 不包含任何和号。

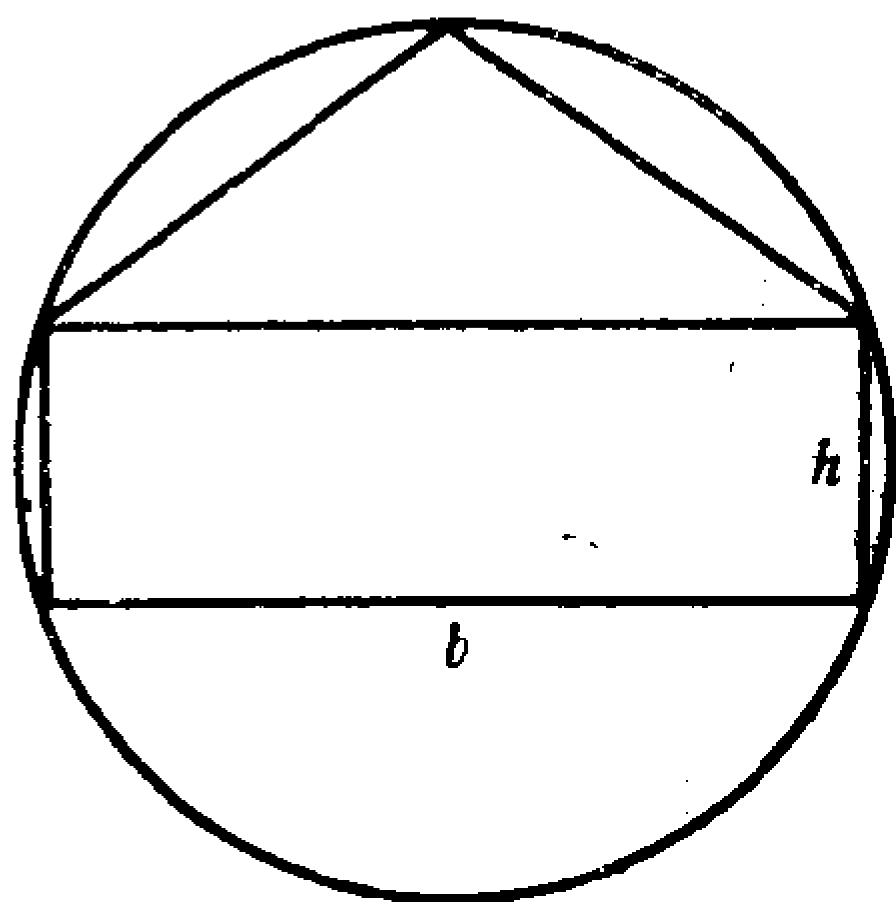



图　1

问题 B-1　在一个半径为1的圆内，有一个底为 $b$，高为 $h$ 的内接矩形，有一个底为 $b$ 的内接等腰三角形，如图1所示。

求 $h$ 的值，使矩形和三角形的面积相等。

问题B-2　设复数 $x,y$ 和 $z$ 满足下面的联立方程

$$x(x-1)+2yz=y(y-1)+2xz=z(z-1)+2xy, \qquad (1)$$

证明：有序的三元数组 $T=(x-y,y-z,z-x)$ 只有有限种可能，请列出所有这样的有序数组 $T$。

**问题B-3** 设$\Gamma$由 $x$ 的全体整系数多项式组成。设$f,g\in\Gamma$，$m$是一个正整数，当 $f-g$ 的每项系数都是$m$的整倍数时，我们说

$$f\equiv g(\mathrm{mod}\ m).$$

令 $n$ 和 $p$ 是正整数，$p$ 是素数。给定$f,g,h,r$和 $s$ 均$\in\Gamma$，满足

$$rf+sg\equiv 1(\mathrm{mod}\ p) \text{ 和 } fg\equiv h(\mathrm{mod}\ p).$$

证明：在$\Gamma$内存在 $F$ 和 $G$，满足

$$F\equiv f(\mathrm{mod}\ p),\qquad G\equiv g(\mathrm{mod}\ p)$$

和

$$FG\equiv h(\mathrm{mod}\ p^n).$$

**问题B-4** 设 $r$ 为一个正实数，令

$$G(r)=\min\{|r-\sqrt{m^2+2n^2}|\,|\,m,n\text{为整数}\}.$$

证明或否定下述论断：

$$\lim_{r\to\infty}G(r)\text{存在且等于零。}$$

**问题B-5** 设$f(x,y,z)=x^2+y^2+z^2+xyz$。设实系数多项式$p(x,y,z)$，$q(x,y,z)$，$r(x,y,z)$满足

$$f(p(x,y,z),\ q(x,y,z),\ r(x,y,z))=f(x,y,z). \tag{**}$$

证明或否定下述论断：系列 $p,q,r$ 由$\pm x,\pm y,\pm z$ 的某个置换组成，其中负号的个数是 0 或 2。

**问题B-6** 设 $A$、$B$、$C$、$D$是 $n\times n$ 矩阵，它们的元素都在

域$F$中，满足以下条件：$AB^t$和$CD^t$是对称的；$AD^t-BC^t=I$。$I$是$n\times n$单位矩阵。若$M$是$n\times n$矩阵，$M^t$表示$M$的转置。

证明：$A^tD-C^tB=I$。

## 解　答

在每个题目的题号后，有一个12维向量$(n_{10},n_9,\cdots,n_0,n_{-1})$。其中$n_i(0\leqslant i\leqslant 10)$表示考得最好的201个参赛者中，此题得$i$分的人数，$n_{-1}$表示没作此题的人数。

在下面解答中，凡属译者增加的注释，均用＊＊……＊＊表示。

**A-1**(152,23,10,7,0,0,0,2,2,3,1,1)

**解**　条件$x^4+36\leqslant 13x^2$等价于

$$(x-3)(x-2)(x+2)(x+3)\leqslant 0,$$

因此

$$S=[-3,-2]\cup[2,3].$$

$f'(x)=3x^2-3$。当$x\in S$，有$f'(x)>0$，$f(x)$在$[-3,-2]$及$[2,3]$上都是增函数。因此

$$\max_{x\in S} f(x)=\max\{f(-2),f(3)\}=18.$$

**A-2**(155,0,0,0,0,0,0,0,0,0,33,13)

**解**　记$I\triangleq\left[\dfrac{10^{20000}}{10^{100}+3}\right]$。注意到$\dfrac{3^{200}}{10^{100}+3}<1$，

所以

$$I=\frac{10^{20000}-3^{200}}{10^{100}+3}.$$

$**$记$a=10^{100}$，有

$$I=a^{199}-a^{198}\cdot 3+a^{197}\cdot 3^2-\cdots+a\cdot 3^{198}-3^{199}**$$

$$I\equiv -3^{199}(\bmod 10)\equiv -3^3(3^4)^{49}(\bmod 10)$$

$$\equiv -7(\bmod 10)\equiv 3(\bmod 10).$$

因此 $I$ 的个位数字是 3 .

A-3(53,6,15,1,0,0,0,1,12,1,26,86)

**解** 利用

$$\cot(\alpha-\beta)=\frac{\cot\alpha\cdot\cot\beta+1}{\cot\beta-\cot\alpha},$$

设 $\cot\alpha=n$，$\cot\beta=n+1$，则 $\cot(\alpha-\beta)=n^2+n+1$，得到

$$\operatorname{Arc}\cot(n^2+n+1)=\operatorname{Arc}\cot n-\operatorname{Arc}\cot(n+1).$$

$$\sum_{n=0}^{\infty}\operatorname{Arc}\cot(n^2+n+1)$$

$$=\lim_{n\to\infty}\sum_{i=0}^{n}\operatorname{Arc}\cot(i^2+i+1)$$

$$=\lim_{n\to\infty}(\operatorname{Arc}\cot 0-\operatorname{Arc}\cot(n+1))$$

$$=\pi/2.$$

A-4(21,3,4,4,5,7,0,6,3,7,23,118)

我们先证

引理 任意的$c_i$和$d_j(i,j=1,2,\cdots,n)$，令$a_{ij}=c_i+d_j$，则导出唯一的一个矩阵$(a_{ij})$，满足条件(b)。反之，如果一个$n\times n$ 矩阵$(a_{ij})$满足条件(b)，则存在唯一一组数 $c_1=0$，$c_2,\cdots,c_n$和$d_1,d_2,\cdots,d_n$使得 $a_{ij}=c_i+d_j$。

**证明** 如果$a_{ij}=c_i+d_j$，则$A$的任一横截的和为

$$\sum_{i=1}^{n} c_i + \sum_{j=1}^{n} d_j,$$

因此(b)满足。

反之，如果$(a_{ij})$满足条件(b)，定义$d_j = a_{1j}$和$c_i = a_{i1} - d_1 = a_{i1} - a_{11}$[注 1]，有$c_1 = 0$。由条件(b)知道

$$a_{ij} + a_{11} = a_{i1} + a_{1j},$$

因此

$$a_{ij} = a_{i1} + a_{1j} - a_{11} = c_i + d_j.$$

所定义的$d_j, c_i$符合要求。

下面讨论$c_i$和$d_j$的唯一性。因为$c_1 = 0$，$a_{1j} = c_1 + d_j$，必有$d_j = a_{1j}$。因为$a_{i1} = c_i + d_1$，必有$c_i = a_{i1} - d_1$。引理证毕。

问题变为讨论$2n$维有序数组$(c_1 = 0, c_2, \cdots, c_n, d_1, d_2, \cdots, d_n)$，由条件(a)知，对所有的$i, j$，它满足$c_i + d_j = 0, \pm 1$。$f(n)$就是上述不同的有序数组的个数。

注意到$c_1 = 0$，因此$d_j \in \{0, 1, -1\}$，$c_i \in \{0, \pm 1, \pm 2\}$。下面我们分八种情形讨论[注 2]：

| $c_i$的取值范围 | $d_j$可能的取值 | $(c_2, \cdots, c_n)$的个数 | $(d_1, \cdots, d_n)$的个数 | 乘　积 |
|---|---|---|---|---|
| 0 | 0, −1, 1 | 1 | $3^n$ | $3^n$ |
| 0, −2 | 1 | $2^{n-1}-1$ | 1 | $2^{n-1}-1$ |
| 0, −1 | 0, 1 | $2^{n-1}-1$ | $2^n$ | $\frac{1}{2}4^n - 2^n$ |
| 0, −1, −2 | 1 | $3^{n-1}-2^n+1$ | 1 | $3^{n-1}-2^n+1$ |
| 0, 2 | −1 | $2^{n-1}-1$ | 1 | $2^{n-1}-1$ |
| 0, 1 | 0, −1 | $2^{n-1}-1$ | $2^n$ | $\frac{1}{2}4^n - 2^n$ |
| 0, 1, 2 | −1 | $3^{n-1}-2^n+1$ | 1 | $3^{n-1}-2^n+1$ |
| 0, −1, 1 | 0 | $3^{n-1}-2^n+1$ | 1 | $3^{n-1}-2^n+1$ |

**对表中第三列有必要作如下说明：在我们填写每一行时，都必须排除前面各行已计数过的情形。例如第二行，$(c_2,\cdots,c_n)$中，每个$c_i$可取0或$-2$，有$2^{n-1}$种可能，但应除去第一行已出现过的全为0的一种，故应填写$2^{n-1}-1$。再看第四行，$(c_2,\cdots,c_n)$中，每个$c_i$可取$0,-2,-1$，共有$3^{n-1}$种可能，但应除去前三行中已出现过的情况，故应填写$3^{n-1}-2(2^{n-1}-1)-1=3^{n-1}-2^n+1$。

第四列填写的是，当$(c_2,\cdots,c_n)$确定后，$(d_1,\cdots,d_n)$有多少种可能情形。

最后一列是第三、四列的乘积。**

把最后一列加起来，给出

$$f(n)=4^n+2\times3^n-4\times2^n+1.$$

注1　原文为$c_i=a_{i1}-a_{1j}$，有误。

注2　表格中的第1至第8行，在原文中依次为第1,2,7,3,4,6,5,8行，原文的安排不便于解释第三列。

A-5(13,4,0,0,0,0,0,1,0,2,39,142)

**证明**　注意到$c_{ji}=-c_{ij}$（对所有的$i,j$）。令$h_i=\frac{1}{2}\sum_j c_{ij}x_j$，有$\partial h_i/\partial x_j=\frac{1}{2}c_{ij}$。因此

$$\frac{\partial h_i}{\partial x_j}-\frac{\partial h_j}{\partial x_i}=\frac{1}{2}c_{ij}-\frac{1}{2}c_{ji}=c_{ij}=\frac{\partial f_i}{\partial x_j}-\frac{\partial f_j}{\partial x_i},$$

对所有的$i,j$有

$$\frac{\partial(h_i-f_i)}{\partial x_j}=\frac{\partial(h_j-f_j)}{\partial x_i}.$$

因此$(h_1-f_1,\cdots,h_n-f_n)$是一个梯度，(**由微积分中斯托克斯积分理论**)可断定存在一个函数$g$,使$\frac{\partial g}{\partial x_i}=(h_i-f_i)$，即$f_i+\partial g/\partial x_i=h_i$是线性函数。

A-6(1,4,1,1,0,1,0,0,6,4,64,119)

解　记$(b)_j=b(b-1)\cdots(b-j+1)$。对题中(*)式作$j$次微分$(0\leqslant j\leqslant n)$，然后置$x=1$，得到$(n+1)$个方程

$$
\begin{aligned}
0 &= 1+\sum a_i,\\
0 &= \sum a_i b_i,\\
0 &= \sum a_i (b_i)_2,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
0 &= \sum a_i (b_i)_{n-1},\\
n!f(1) &= \sum a_i (b_i)_n.
\end{aligned}
$$

用克莱姆法则解前$n$个方程，然后把解出的$a_1,\cdots,a_n$代入最后一个方程，我们得到

$$
n!f(1)=\frac{\sum_{i=1}^{n}(b_i)_n(-1)^i\begin{vmatrix} b_1 & \cdots & \hat{b}_i & \cdots & b_n\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ (b_1)_{n-1} & \cdots & (\hat{b}_i)_{n-1} & \cdots & (b_n)_{n-1}\end{vmatrix}}{\begin{vmatrix} 1 & 1 & \cdots & 1\\ b_1 & b_2 & \cdots & b_n\\ \cdots & \cdots & \cdots & \cdots\\ (b_1)_{n-1} & (b_2)_{n-1} & \cdots & (b_n)_{n-1}\end{vmatrix}}
$$

此处，上标^表示实际不存在的项。

分母$D$是$b_1, b_2, \cdots, b_n$的$\binom{n}{2}$次齐次多项式，如令$b_i = b_j$ $(i \neq j)$，则$D=0$，可知$D$有因式$(b_i - b_j)$，因此

$$D = c\prod_{i>j}(b_i - b_j),\quad c\text{为常数，}$$

注意$b_n^{n-1}b_{n-1}^{n-2}\cdots b_3^2 b_2^1$这一项在$D$及$\prod_{i>j}(b_i - b_j)$中的系数均为1，因此$D = \prod_{i>j}(b_i - b_j)$①。

我们也能用另一种方法讨论分母$D$，我们对$D$作行的初等变换，把它变为Vandermonde(范德蒙)行列式，同样得到$D = \prod_{i>j}(b_i - b_j)$。

再讨论分子，如令$b_i = b_j (i \neq j)$，则分子的$\sum$号中只有两项不等于零，但这两项大小相等，符号相反．可知分子可被$(b_i - b_j)$整除，于是分子可被分母整除，因此$n!f(1)$是一个$b_1, \cdots, b_n$的多项式。注意到分子的$\sum$号中每项的次数比分母高$n$次，因此多项式$n!f(1)$的次数$\leqslant n$。

令$b_i = 0$，此时分母$\neq 0$，但是分子的每项是零，因此多项式$n!f(1)$被$b_i (1 \leqslant i \leqslant n)$整除，$n!f(1) = kb_1 b_2 \cdots b_n$，$k$为常数。

* * 只要$\{a_i | a_i$为实数$, 1 \leqslant i \leqslant n\}$，$\{b_j | b_j$为不同正整数，$1 \leqslant j \leqslant n\}$和$f(x)$使题中(*)式成立，导出的$k$值均相同。显然，当$\{b_j = j,\ 1 \leqslant j \leqslant n\}$时，取$f(x) \equiv 1$，存在一组$a_i$使

① 原文为$\prod_{i<j}(b_i - b_j)$，有误。——译注

(*)式成立，此时有$k=1$。**因此

$$f(1)=b_1b_2\cdots b_n/n!。$$

B-1(183,3,7,0,0,0,0,0,4,2,1,1)

**解** 三角形的高是$\frac{1}{2}(2-h)$。由面积相等得出

$$h=(\text{三角形的高})/2=(2-h)/4,$$

因此

$$h=\frac{2}{5}.$$

B-2 (123,31,16,3,0,0,0,0,16,5,2,5)

**证明** 联立方程组(1)等价于

$$0=(x-y)(x+y-1-2z)=(y-z)(y+z-1-2x)$$
$$=(z-x)(z+x-1-2y),$$

如果$x$、$y$,$z$中任意两个均不相等，得到

$$x+y-1-2z=y+z-1-2x=z+x-1-2y=0.$$

把上述三个方程加在一起，得到矛盾式$-3=0$。因此，$x,y,z$中至少有两个相等。

设$x=y$，$y\neq z$，则$z=x+1$。在这种情形有$x-y=0$，$y-z=-1$，$z-x=1$。可类似讨论$y=z,z\neq x$和$z=x,x\neq y$的情形。

因此$T=(x-y,y-z,z-x)$只能取下述值

$$(0,0,0),\ (0,-1,1),\ (1,0,-1)\text{和}(-1,1,0).$$

容易看出，上述四种取值都符合题目要求。

B-3 (26,5,4,1,0,1,0,4,3,5,33,119)

**证明** 对 $n$ 用归纳法。当 $n=1$ 时，取 $F_1=f, G_1=g$，就符合要求。假设 $n=k$，已有 $F_k, G_k \in \Gamma$，满足 $F_k \equiv f \pmod p$，$G_k \equiv g \pmod p$，且 $F_k G_k \equiv h \pmod{p^k}$。

下面讨论 $n=k+1$ 的情形。记 $h-F_kG_k=tp^k$，其中 $t\in\Gamma$。令 $F_{k+1}=F_k+stp^k$ 和 $G_{k+1}=G_k+rtp^k$。则

$$F_{k+1}\equiv F_k\equiv f \pmod p,\quad G_{k+1}\equiv G_k\equiv g \pmod p,$$

并且

$$\begin{aligned}F_{k+1}G_{k+1}&=F_kG_k+tp^k(rF_k+sG_k)+rst^2p^{2k}\\&\equiv F_kG_k+tp^k(rF_k+sG_k) \pmod{p^{k+1}}.\end{aligned}$$

由假设条件知

$$rF_k+sG_k\equiv rf+sg\equiv 1 \pmod p,$$

于是我们可设 $rF_k+sG_k=1+qp$，其中 $q\in\Gamma$。因此

$$\begin{aligned}F_{k+1}G_{k+1}&\equiv F_kG_k+tp^k(1+qp) \pmod{p^{k+1}}\\&\equiv F_kG_k+tp^k \pmod{p^{k+1}}\\&\equiv h \pmod{p^{k+1}}.\end{aligned}$$

由归纳法，我们完成了证明（题目的结论成立，与 $p$ 是否素数无关，同时与多项式的变元个数也无关）。

B-4 (22,8,6,6,0,0,0,0,4,7,59,89)

**证明** 令 $m$ 是满足 $r^2\geqslant m^2$ 的最大的非负整数，令 $n$ 是满足 $(r^2-m^2)/2\geqslant n^2$ 的最大的非负整数。于是 $r^2-m^2<(m+1)^2-m^2=2m+1\leqslant 2r+1$，及

$$0\leqslant\frac{r^2-m^2}{2}-n^2<(n+1)^2-n^2=2n+1\leqslant 2\sqrt{\frac{r^2-m^2}{2}}+1$$

$$<2\sqrt{\frac{2r+1}{2}}+1=\sqrt{2(2r+1)}+1.$$

即 $r^2-m^2-2n^2<2\sqrt{2\sqrt{2r+1}}+2$。因为

$$r^2-m^2-2n^2=(r-\sqrt{m^2+2n^2})\cdot(r+\sqrt{m^2+2n^2}),$$

$$0\leqslant r-\sqrt{m^2+2n^2}=\frac{r^2-m^2-2n^2}{r+\sqrt{m^2+2n^2}}$$

$$<\frac{2\sqrt{2\sqrt{2r+1}}+2}{r},$$

当 $r\to\infty$ 时，$r-\sqrt{m^2+2n^2}\to 0$。由

$$0\leqslant G(r)\leqslant r-\sqrt{m^2+2n^2},$$

得出

$$\lim_{r\to\infty} G(r)=0.$$

**B-5** (10,0,0,0,0,0,0,0,0,0,58,133)

论断是不成立的。

注意下述事实：若 $(p,q,r)$ 满足(＊)式，则 $(p,q,-r-pq)$ 也满足(＊)式。因此，取 $p=x$，$q=y$，$r=-z-xy$，可满足(＊)式，就是一个反例。

**B-6** (3,0,0,0,0,0,0,0,0,0,32,166)

**证明** 问题的条件是

(i) $AB^t=(AB^t)^t=BA^t$,

(ii) $CD^t=(CD^t)^t=DC^t$,

(iii) $AD^t-BC^t=I$.

由条件(i)推出 $BA^t-AB^t=0$ ($n\times n$ 零矩阵)，由条件(ii)推出 $CD^t-DC^t=0$，由条件(iii)的转置推出 $DA^t-CB^t=I^t=I$。因此我们有

$$\begin{pmatrix}A & B\\ C & D\end{pmatrix}\begin{pmatrix}D^t & -B^t\\ -C^t & A^t\end{pmatrix}=\begin{pmatrix}AD^t-BC^t & -AB^t+BA^t\\ CD^t-DC^t & -CB^t+DA^t\end{pmatrix}$$

$$=\begin{pmatrix} I & O \\ O & I \end{pmatrix},$$

由此推出

$$\begin{pmatrix} D^t & -B^t \\ -C^t & A^t \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix}=\begin{pmatrix} I & O \\ O & I \end{pmatrix},$$

用分块乘法计算右下角元素，得到

$$-C^tB+A^tD=I.$$

证毕。

（杨晓兰译）

# 第三十届国际数学奥林匹克竞赛试题及解答

徐明曜　罗华章　等

**编者按**　第三十届国际奥林匹克数学竞赛于1989年7月18,19日两天在联邦德国的Braunschweig举行。我国取得了四枚金奖、两枚银奖，获团体总分第一的好成绩。来自四川永川中学的朴实的罗华章同学六题全对，获得整个比赛的金奖第一名，他已进入北京大学数学系学习。我们请徐明曜、潘承彪、杨晚兰和王鲁燕编写了一份解答，同时请罗华章同学提供了他参赛时的解答（我们作了若干文字上的修饰），供读者参考。

## 试　题

1．求证：集合$\{1,2,\cdots,1989\}$可以分为117个互不相交的子集$A_i(i=1,2,\cdots,117)$，使得

(1) 每个$A_i$含有17个元素；

(2) 每个$A_i$中各元素之和相同。

2．锐角三角形$ABC$中，$A$角的等分线与三角形的外接圆交于另一点$A_1$。点$B_1,C_1$与此类似。直线$AA_1$与$B,C$两角的外角等分线相交于$A_0$。点$B_0,C_0$与此类似。

求证：

(1) 三角形 $A_0B_0C_0$ 的面积是六边形 $AC_1BA_1CB_1$ 面积的二倍；

(2) 三角形 $A_0B_0C_0$ 的面积至少是三角形 $ABC$ 面积的四倍。

3. 设 $n$ 和 $k$ 是正整数，$S$ 是平面上 $n$ 个点的集合，满足：

(1) $S$ 中任何三点不共线；

(2) 对 $S$ 中的每一个点 $P$，$S$ 中至少存在 $k$ 个点与 $P$ 距离相等。

求证：

$$k<\frac{1}{2}+\sqrt{2n}.$$

4. 设 $ABCD$ 是一个凸四边形，它的三个边 $AB,AD,BC$ 满足 $AB=AD+BC$。四边形内，距离 $CD$ 为 $h$ 的地方有一点 $P$，使得 $AP=h+AD,BP=h+BC$。求证：

$$\frac{1}{\sqrt{h}}\geqslant\frac{1}{\sqrt{AD}}+\frac{1}{\sqrt{BC}}.$$

5. 求证：对任何正整数 $n$，存在 $n$ 个相继的正整数，它们都不是素数的整数幂。

6. 设 $n$ 是正整数。我们说集合$\{1,2,\cdots,2n\}$的一个排列$(x_1,x_2,\cdots,x_{2n})$具有性质 $P$，如果在$\{1,2,\cdots,2n-1\}$当中至少有一个 $i$ 使$|x_i-x_{i+1}|=n$成立。求证：对于任何 $n$，具有性质 $P$ 的排列比不具有性质 $P$ 的排列个数多。

## （一）徐明曜等同志的解答

**第 1 题解答** 我们把问题提得更一般些．设 $s,n$ 是两个正整数．证明集合 $A=\{1,2,\cdots,sn\}$ 可以分为 $n$ 个互不相交的子集 $A_1,\cdots,A_n$，使得每个集合含有 $s$ 个元素，且每个集合中各元素之和相同．本题就是 $s=17$，$n=117$．

以 $S(A)$ 和 $S(A_i)$ 分别表集合 $A$ 和 $A_i$ 中各元素之和．我们有

$$S(A)=sn(sn+1)/2. \tag{1}$$

由于这里要求 $S(A_1)=\cdots=S(A_n)$，所以

$$S(A_i)=s(sn+1)/2,\quad i=1,2,\cdots,n. \tag{2}$$

由式(2)看出，当 $s$ 为奇数、$n$ 为偶数时，这种分法是不可能的．所以我们只要讨论两种情形：(i) $s$ 为偶数；(ii) $s,n$ 均为奇数．我们的题目属于情形(ii)．情形(i)是很容易解决的，(ii)要困难些．

情形(i) 我们把这 $sn$ 个数用下面的方法自小至大依次排成一个 $s$ 行、每行有 $n$ 个数的表，即 $s$ 行 $n$ 列的表：第 1 行从 1 开始自左向右依次排到 $n$，然后排第 2 行，要注意的是把 $n+1$ 排在 $n$ 的下面（即在同一列），且和第 1 行排列的方向相反，自右向左排到 $2n$．一般的，排好第 $j$ 行的最后一个数 $jn$ 后，再排第 $j+1$ 行，把 $jn+1$ 排在 $jn$ 的下面，且和第 $j$ 行的排列方向相反，依次排到 $(j+1)n$．表 1 就给出了当 $s$ 是偶数时这样的排列．

由排列的方法立即看出：任意相邻两行在同一列上的两数之和都相同．由于现在 $s$ 是偶数，所以，第 1,2 行，第 3,

表 1

| $A_1$ | $A_2$ | $A_3$ | … | $A_{n-2}$ | $A_{n-1}$ | $A_n$ |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | … | $n-2$ | $n-1$ | $n$ |
| $2n$ | $2n-1$ | $2n-2$ | … | $n+3$ | $n+2$ | $n+1$ |
| $2n+1$ | $2n+2$ | $2n+3$ | … | $3n-2$ | $3n-1$ | $3n$ |
| $4n$ | $4n-1$ | $4n-2$ | … | $3n+3$ | $3n+2$ | $3n+1$ |
| … | … | … | … | … | … | … |
| $(s-4)n+1$ | $(s-4)n+2$ | $(s-4)n+3$ | … | $(s-3)n-2$ | $(s-3)n-1$ | $(s-3)n$ |
| $(s-2)n$ | $(s-2)n-1$ | $(s-2)n-2$ | … | $(s-3)n+3$ | $(s-3)n+2$ | $(s-3)n+1$ |
| $(s-2)n+1$ | $(s-2)n+2$ | $(s-2)n+3$ | … | $(s-1)n-2$ | $(s-1)n-1$ | $(s-1)n$ |
| $sn$ | $sn-1$ | $sn-2$ | … | $(s-1)n+3$ | $(s-1)n+2$ | $(s-1)n+1$ |

4 行，…，第 $s-1,s$ 行都是相邻的两行，且恰好把这 $s$ 行分完。因此推出：这表中每一列中的元素之和都相同。这样，只要把第 $i$ 列中的数组成的集合取作子集 $A_i$，就证明了所要的结论。

当 $s$ 是奇数时虽然也可以这样列表，但不能以相邻两行为一组把这 $s$ 行分完，所以这样的方法不能解决 $s$ 是奇数的情形。

应该指出，把集合 $A$ 改为是由任意相邻的 $sn$ 个整数组成的集合，即取 $A=\{a+1,a+2,\cdots,a+sn\}$时（这里 $a$ 是任意给定的整数），当 $s$ 为偶数时结论也成立。事实上只要在表中每个数加上 $a$ 即可推出。这一点在证情形(ii)时要用到。

情形(ii)　我们只要讨论 $s=3$ 的情形。因为当奇数 $s>3$ 时可把集合 $A=\{1,2,\cdots,sn\}$分为两个集合：$B=\{1,2,\cdots,3n\}$，

$C=\{3n+1,3n+2,\cdots,3n+(s-3)n\}$。集合$C$有$(s-3)n$个数，$s-3$是偶数，由情形(i)最后的说明知，它可以分为互不相交的$n$个子集$C_1,\cdots,C_n$，每个子集有$s-3$个数，且每个子集中的数之和都相同．如果我们把集合$B$也分成了互不相交的$n$个子集$B_1,\cdots,B_n$，每个子集有3个数，且每个子集中的数之和都相同，那末，取$A_i$为$B_i$与$C_i$的和集$(i=1,2,\cdots,n)$，这样得到的$A_1,\cdots,A_n$就满足要求。

现在来考虑如何分集合$B$，这里$n$是奇数。通过对具体数值例子（如取$n=3,5,7$等）试验，使我们会想到这样来分集合$B$：先把集合$B$分为子集$E_1=\{1,2,\cdots,2n\}$，和$E_2=\{2n+1,\cdots,3n\}$；然后把集合$E_1$分为$n$个互不相交的子集$E_{11},\cdots,E_{1n}$，每个$E_{1i}$有两个数，且$S(E_{11}),S(E_{12}),\cdots,S(E_{1n})$是$n$个相邻整数（这里$S(E_{1i})$表集合$E_{1i}$的数之和）。如果这样的分法是可能的，不妨假定

$$S(E_{1\,i+1})=S(E_{1i})+1, \tag{3}$$

那么，取子集$B_i$为子集$E_{1i}$再添加$E_2$中的一个数$3n-(i-1)$所组成，这样集合$B$就分成了互不相交的$n$个集合$B_1,\cdots,B_n$，显见它们满足要求。

那末，如何来实现集合$E_1$的满足式(3)的这种分法呢？先来看一个具体例子。$n=5$时，$E_1=\{1,2,\cdots,10\}$可这样来分：

| $E_{11}$ | $E_{14}$ | $E_{12}$ | $E_{15}$ | $E_{13}$ |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |
| 8 | 10 | 7 | 9 | 6 |

也可以这样分：

| | | | | |
|---|---|---|---|---|
| $E_{13}$ | $E_{11}$ | $E_{14}$ | $E_{12}$ | $E_{15}$ |
| 1 | 2 | 3 | 4 | 5 |
| 10 | 7 | 9 | 6 | 8 |

由此得到启发，对一般的集合 $E_1$，按第一种分法可这样来分：先自左至右依次写下 $1,2,\cdots,n$；然后这样写第 2 行：在 $n$ 下面写 $n+1$，再自右向左一隔一地依次写下 $n+2,\cdots$，直到在 1 的下面写 $n+(n+1)/2$（注意 $n$ 是奇数），再回过来从 $n-1$ 下面的第 2 行的空格开始，自右向左依次在这些空格中写完余下的数。下表就给出了这样的分法：

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $G_1$ | $G_2$ | $G_3$ | $G_4$ | $\cdots$ | $G_{n-3}$ | $G_{n-2}$ | $G_{n-1}$ | $G_n$ |
| 1 | 2 | 3 | 4 | $\cdots$ | $n-3$ | $n-2$ | $n-1$ | $n$ |
| $n+\frac{n+1}{2}$ | $\underline{2n}$ | $n+\frac{n-1}{2}$ | $\underline{2n-1}$ | $\cdots$ | $\underline{n+\frac{n+5}{2}}$ | $n+2$ | $\underline{n+\frac{n+3}{2}}$ | $n+1$ |

若以 $G_i$ 记上表中第 $i$ 列数组成的集合，那末，依指标大小依次先写奇指标集合，写完后再写偶指标集合，得到：$G_1$，$G_3,\cdots,G_n,G_2,G_4,\cdots,G_{n-1}$。不改变这一列集合的次序，改写为 $E_{11},E_{12},\cdots,E_{1n-1},E_{1n}$，这就是我们所要的集合 $E_1$ 的分法，且满足式(3)。因为，$S(G_1),S(G_3),\cdots,S(G_n)$ 是递增的相邻整数，$S(G_2),S(G_4),\cdots,S(G_{n-1})$ 也是递增的相邻整数，这两点由表的写法可直接看出，再由写法知，$S(G_n)=2n+1$，$S(G_2)=2n+2$，也是相邻的。

综上所述，在情形(ii)我们也可得到所要的结论。证毕。

最后，我们要指出两点：(a) 由两种情形的讨论可看出，除了显然情形外，集合 $A$ 分为两两不相交的这种子集 $A_1,\cdots$，$A_n$ 的分法不是唯一的，讨论分法的个数应该是一个有趣的问题，可能并不容易。

(b) 这道题目出现 1989 当然同今年是 1989 年有关。但更有意义的是大家知道这样一个故事：伟大的数学家 Gauss 在八岁时，用巧妙的算法迅速算出 $1+2+\cdots+100$ 的和等于 5050。显然情形(i)中列表的方法实际上就是他的算法。而情形(ii)中分解集合 $E_1$ 的方法只不过是 Gauss 算法的一个推广。今年的比赛是在 Gauss 的诞生地联邦德国的 Braunschweig 举行的，因此，把本题选作本届比赛的第 1 题是十分恰当、十分有纪念意义的。

**第 2 题解答** 记$\triangle ABC$的内心为 $I$，外心为 $O$，外接圆半径为 $R$，三内角依次为 $\alpha,\beta,\gamma$。记六边形 $AC_1BA_1CB_1$ 的面积为 $S^*$（见图 1）。

注意到 $S_{\triangle A_1OB}=\frac{1}{2}R^2\sin\alpha$， 以及有关 $\triangle BOC_1$，$\triangle C_1OA$，$\triangle AOB_1$，$\triangle B_1OC$ 和$\triangle COA_1$的面积的类似等式，得出

$$S^*=\frac{1}{2}R^2(2\sin\alpha+2\sin\beta+2\sin\gamma).$$

注意到 $S_{\triangle COB}=\frac{1}{2}R^2\sin 2\alpha$，以及有关$\triangle BOA$和$\triangle AOC$的面积的类似等式，得出

$$S_{\triangle ABC}=\frac{1}{2}R^2(\sin 2\alpha+\sin 2\beta+\sin 2\gamma).$$

化简

$$\begin{aligned}\sin 2\alpha+\sin 2\beta&=2\sin(\alpha+\beta)\cdot\cos(\alpha-\beta)\\&=2\sin\gamma\cdot\cos(\alpha-\beta).\end{aligned}$$

于是，

$$S_{\triangle ABC}=\frac{1}{2}R^2[\sin\gamma\cdot\cos(\alpha-\beta)+\sin\alpha\cdot\cos(\beta-\gamma)$$
$$+\sin\beta\cdot\cos(\alpha-\gamma)],$$

因此

$$S^*\geqslant 2S_{\triangle ABC}.$$

我们看到 $I$ 是$\triangle A_0B_0C_0$ 的垂心。因此，$A_0, C, I, B$ 四点共圆。下面证明 $A_1$ 是圆心。

$$\angle A_1CI=\frac{1}{2}(\overset{\frown}{A_1B}+\overset{\frown}{BC_1})=\frac{1}{2}(\alpha+\gamma),$$

$$\angle A_1IC=\frac{1}{2}(\overset{\frown}{A_1C}+\overset{\frown}{AC_1})=\frac{1}{2}(\alpha+\gamma).$$

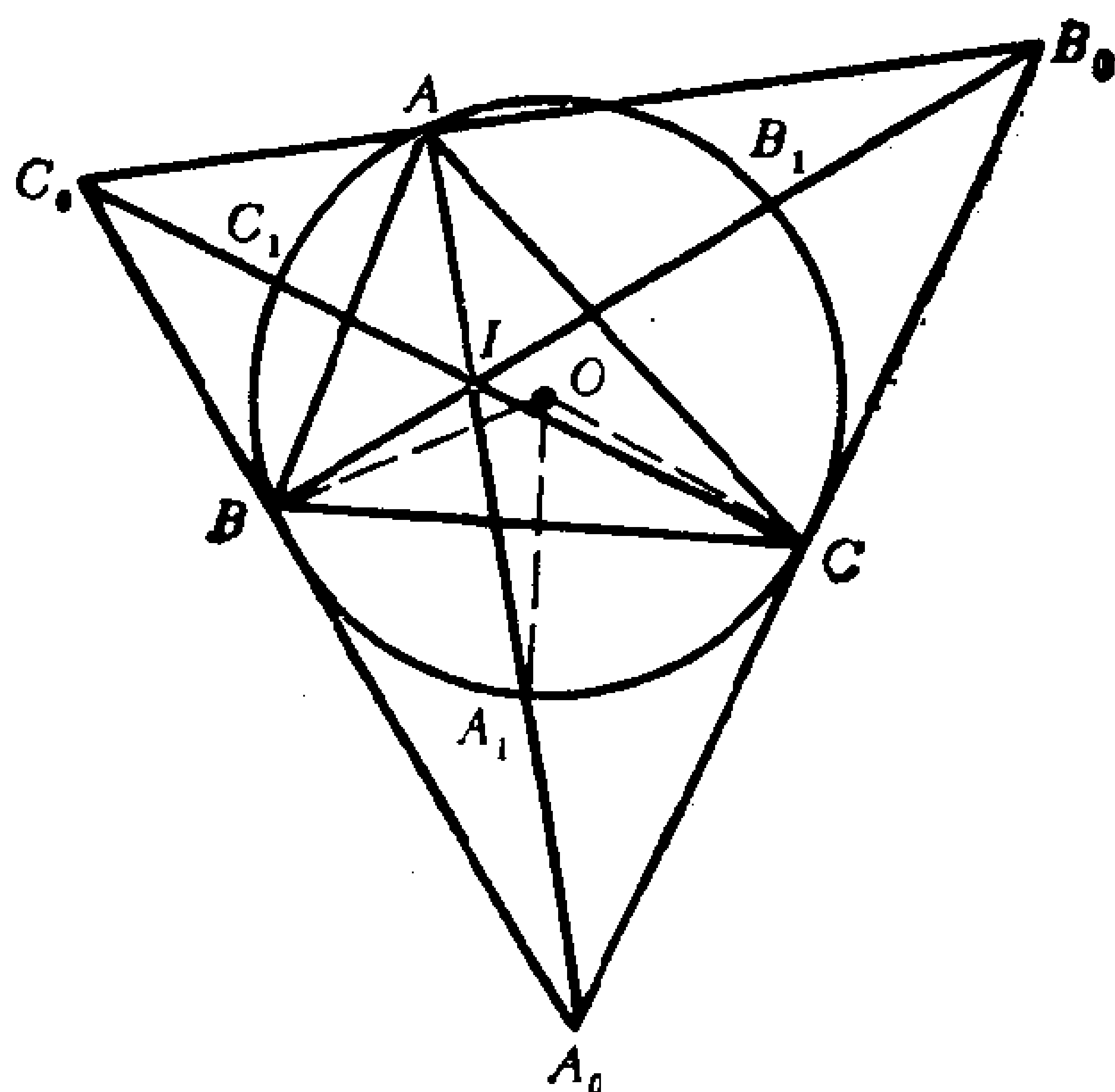



图 1

于是，$A_1C=A_1I$。又有 $A_1C=A_1B$。因此，$A_1$ 是过$B$, $I$,$C$ 三点的圆的圆心，$A_0$ 也在该圆上．因为 $A_1$是$IA_0$的中点，所以$S_{A_0CIB}=2\times S_{A_1CIB}$，再对 $S_{B_0AIC}$ 和 $S_{C_0BIA}$ 作类似

分析。有

$$S_{\triangle A_0B_0C_0}=2S^*,$$

及

$$S_{\triangle A_0B_0C_0}\geqslant 4S_{\triangle ABC}.$$

**第3题解答** 由题意知对每点 $P_i\in S$，存在以 $P_i$ 为圆心的一个圆 $C_i$，使得 $C_i$ 上至少有 $S$ 中的 $k$ 个点。我们对每个点 $P_i$ 都取定这样的圆 $C_i$。因 $S$ 中一共有 $n$ 个点，也一共取定了 $n$ 个圆。

称 $(P_i,C_j)$ 为一个"点圆对"，如果 $P_i\in C_j$。称 $(P_i,C_j,C_k)$，$j\neq k$，为一个"点双圆组"，如果 $P_i\in C_j\cap C_k$。

设过 $P_i$ 点共有 $x_i$ 个圆，又设共有 $M$ 个"点圆对"和 $N$ 个"点双圆组"。则显然有

$$M=\sum_{i=1}^{n}x_i,\qquad N=\sum_{i=1}^{n}\frac{x_i(x_i-1)}{2};$$

又显然有

$$M\geqslant nk,\qquad N\leqslant n(n-1).$$

于是

$$\begin{aligned}2n(n-1)\geqslant 2N&=\sum_{i=1}^{n}(x_i^2-x_i)\\&\geqslant\frac{1}{n}\left(\sum_{i=1}^{n}x_i\right)^2-\sum_{i=1}^{n}x_i\\&=\frac{1}{n}M^2-M=\frac{1}{n}\left(M-\frac{n}{2}\right)^2-\frac{n}{4}\\&\geqslant\frac{1}{n}\left(nk-\frac{n}{2}\right)^2-\frac{n}{4}=nk^2-nk.\end{aligned}$$

由此得

$$2n-2\geqslant k^2-k, \qquad 2n-\frac{7}{4}\geqslant\left(k-\frac{1}{2}\right)^2,$$

即

$$k\leqslant\frac{1}{2}+\sqrt{2n-\frac{7}{4}}<\frac{1}{2}+\sqrt{2n}.$$

(注　此题不用条件(1))

**第 4 题解答**　凸四边形 $ABCD$，以点 $A$(点 $B$)为圆心，以 $a=AD$(以 $b=BC$)为半径作圆．按题意，$AB=AD+BC$，故两圆相切于点 $E$（见图 2）．

先考虑特殊情形．凸四边形 $ABC_0D_0$，边 $C_0D_0$ 恰是圆 $A$ 和圆 $B$ 的公切线．设以 $P_0$ 为圆心，以 $h_0$ 为半径的圆与圆 $A$ 及圆 $B$ 都外切，并且与 $C_0D_0$ 相切于点 $F_0$．

$$C_0D_0^2=(a+b)^2-(a-b)^2=4ab,$$
$$D_0F_0^2=(a+h_0)^2-(a-h_0)^2=4ah_0,$$

类似有

$$C_0F_0^2=4bh_0.$$

由于 $C_0F_0+F_0D_0=C_0D_0$，有

$$2\sqrt{ah_0}+2\sqrt{bh_0}=2\sqrt{ab},$$

即

$$\sqrt{\frac{1}{h_0}}=\sqrt{\frac{1}{a}}+\sqrt{\frac{1}{b}}.$$

再回到一般情形．设边 $CD$ 不是圆 $A$ 和圆 $B$ 的公切线，则 $CD$ 一定在(或通过)四边形 $ABC_0D_0$ 的内部．

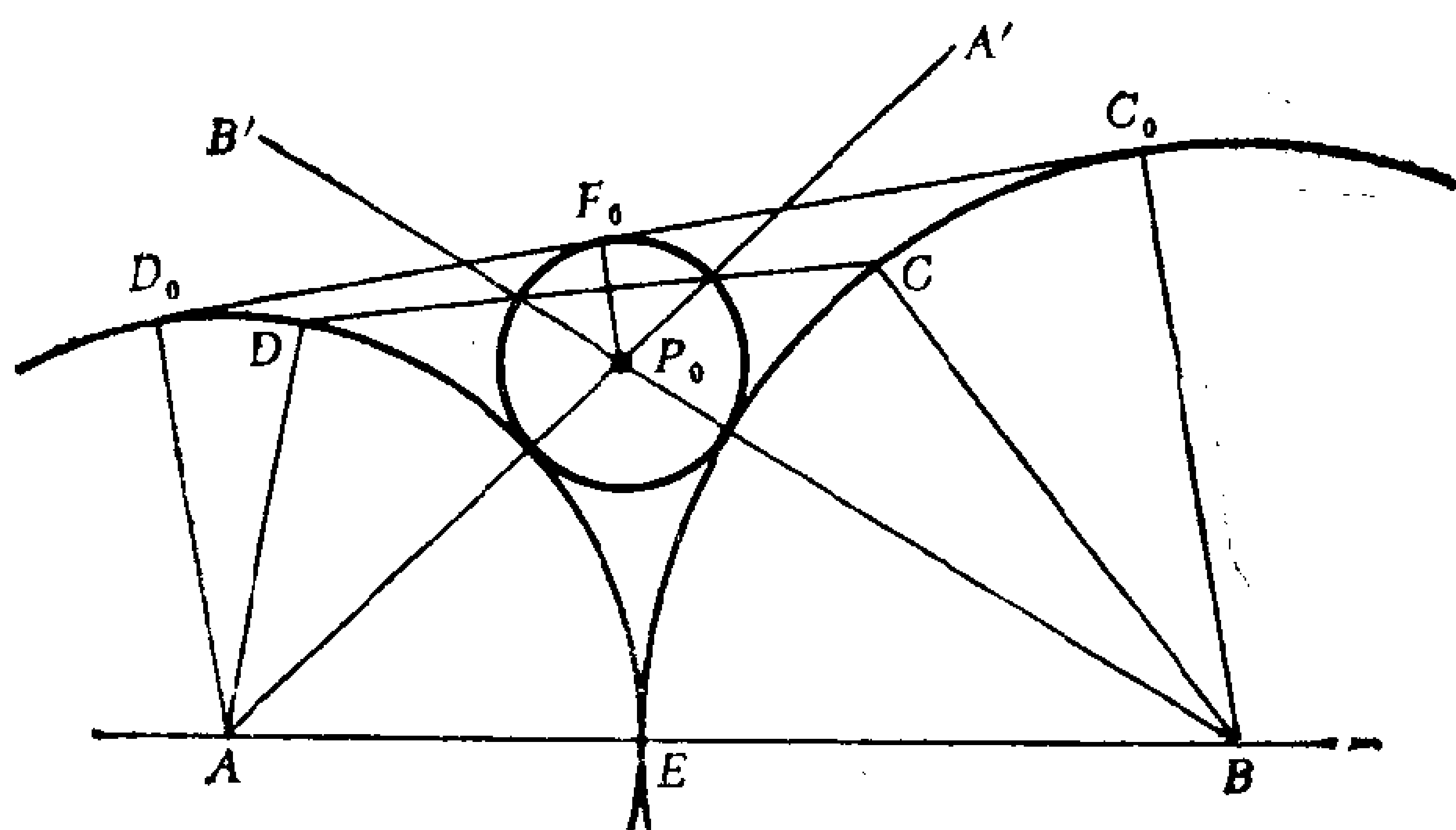



图 2

设点 $P$ 在四边形 $ABC_0D_0$ 的内部，以 $P$ 为心，以 $h$ 为半径的圆与圆 $A$ 及圆 $B$ 都外切。若 $h=h_0$，则点 $P$ 就是 $P_0$；若 $h>h_0$，则点 $P$ 在 $\angle B'P_0A'$ 的内部。这表明，当 $h\geqslant h_0$ 时，圆 $P$ 一定不完全在四边形 $ABCD$ 的内部。因此，若要求圆 $P$ 完全在 $ABCD$ 的内部，且与 $CD$ 相切，必有 $h<h_0$，即

$$\sqrt{\frac{1}{h}}>\sqrt{\frac{1}{a}}+\sqrt{\frac{1}{b}}.$$

**第 5 题解答**　本题即是要找出这样的 $n$ 个相邻正整数，每个都至少有两个不同的素因子。所以这是初等数论的一个整除问题。我们用归纳法来证。$n=1$ 时结论显然成立。假定结论对 $n=k(k\geqslant 1)$ 成立，即存在正整数 $a$ 使得 $k$ 个相邻正整数：

$$a, a+1, \cdots, a+(k-1), \tag{1}$$

每个都有两个不同的素因子，不妨设 $a+i$ 的两个不同的素因

子是$p_i, q_i, i=0,1,\cdots,k-1$。 记$P_k=p_0q_0p_1q_1\cdots p_{k-1}q_{k-1}$。显见，对任意非负整数$t$，$k$个相邻正整数

$$a+tP_k,\ a+tP_k+1,\cdots,a+tP_k+(k-1) \qquad (2)$$

也满足所说的要求，因为$a+tP_k+i$和$a+i$一样也有素因子$p_i, q_i$。我们来证明结论对$n=k+1$也成立，即存在正整数$x$使得$k+1$个相邻正整数。

$$x,x+1,\cdots,x+(k-1),x+k, \qquad (3)$$

每个都有两个不同的素因子。很自然的，希望从形如$a+tP_k$的数中去找这样的$x$，因为这时由归纳假定和前面的说明知，对任意非负正整数$t$，当$x=a+tP_k$时，式(3)的前$k$个正整数都有两个不同的素因子，这样，就只要去决定$t$，使得$x+k=a+tP_k+k$也有两个不同的素因子。下面来求满足这样要求的$t$

假定$x+k$有两个不同的素因子$p_k, q_k$(待定)，就有

$$x+k=sp_kq_k。$$

这样，$s$和$t$满足关系式

$$sp_kq_k-tP_k=a+k。 \qquad (4)$$

这就是熟知的关于变数$s,t$的一次不定方程。我们知道，只要$p_kq_k$和$P_k$互素，不定方程(4)一定有解。由于素数有无穷多个，对已知的$P_k$，一定可以找到不同的素数$p_k, q_k$使$p_kq_k$和$P_k$互素。这样，对取定的这种$p_k, q_k$必有$t_0, s_0$是方程(4)的一组解，且对任意整数$u$，

$$t=t_0+up_kq_k,\quad s=s_0+uP_k$$

也都是方程(4)的解。所以，一定可以取到$u$使$t$为非负整数。因此，结论对$n=k+1$也成立。证毕。

熟悉初等数论的读者，不难看出本题就是解一次同余方程组，利用孙子定理即可证明。也就是对取定的$2n$个不同的素数$p_0,q_0,p_1,q_1,\cdots,p_{n-1},q_{n-1}$，求解一次同余方程组

$$x+i\equiv 0(\bmod p_{i-1}q_{i-1}),\quad i=0,1,\cdots,n-1,$$

的正整数解。

第 6 题解答　在集合$\{1,2,\cdots,2n\}$中，使差的绝对值等于$n$的数对是$\{1,n+1\},\{2,n+2\},\cdots,\{n,2n\}$。因此，具有性质$P$的排列就是这样的排列：它至少有两个相邻数是属于上述$n$个数对之一。我们说一个排列$(x_1,x_2,\cdots,x_{2n})$具有性质$a_j(1\leqslant j\leqslant n)$，如果数对$\{j,n+j\}$中的两个数是这个排列中的相邻数(不计次序)。这样就可用容斥原理来解本题。

以$S$记具有性质$P$的排列个数，以$S_j$记所有具有性质$a_j$的排列的集合及其个数，对取自集合$\{1,2,\cdots,n\}$中的一组数$j_1<j_2<\cdots<j_l$，以$S_{j_1j_2\cdots j_l}$记同时具有性质$a_{j_1},a_{j_2},\cdots,a_{j_l}$的所有排列的集合及其个数。由容斥原理知，

$$\begin{aligned}S=&\sum_{1\leqslant j_1\leqslant n}S_{j_1}-\sum_{1\leqslant j_1<j_2\leqslant n}S_{j_1j_2}+\sum_{1\leqslant j_1<j_2<j_3\leqslant n}S_{j_1j_2j_3}\\&-\cdots+(-1)^{l-1}\sum_{1\leqslant j_1<j_2<\cdots<j_l\leqslant n}S_{j_1j_2\cdots j_l}\\&+\cdots+(-1)^{n-1}S_{12\cdots n}.\end{aligned}\tag{1}$$

由于总的排列数为$(2n)!$，所以本题就是要证明

$$S>(2n)!/2.\tag{2}$$

下面来计算$S_{j_1\cdots j_l}$。 $S_{j_1}$中的每个排列必定是数$j_1$， 和$n+j_1$是这排列中的相邻数(不计次序)，因此，个数$S_{j_1}$就相当于把$j_1$和$n+j_1$看作一个数，再和另外$2l-2$个数一起任意排列的个数，由于$j_1,n+j_1$不计次序，因此

$$S_{j_1}=2\cdot(2n-1)!.$$

进而有

$$\sum_{1\leqslant j_1<n} S_{j_1}=2C_n^1(2n-1)!. \tag{3}$$

同理， $S_{j_1\cdots j_l}$就相当于把$j_1$和$n+j_1$看作一个数，$j_2$和$n+j_2$看作一个数，…，$j_l$和$n+j_l$看作一个数， 再和另外$2n-2l$个数在一起任意排列的个数，由于$j_t,n+j_t$不计次序($t=1,\cdots,l$)，所以

$$S_{j_1\cdots j_l}=2^l\cdot(2n-l)!.$$

进而有

$$\sum_{1<j_1<\cdots<j_l\leqslant n} S_{j_1\cdots j_l}=2^lC_n^l(2n-l)!. \tag{4}$$

由式(1),(3)及(4)推出

$$\begin{aligned}S&=2C_n^1(2n-1)!-2^2C_n^2(2n-2)!+\cdots\\&\quad+(-1)^{l-1}2^lC_n^l(2n-l)!\\&\quad+\cdots+(-1)^{n-1}2^nC_n^n n!\\&=2\cdot(2n-1)!\left\{C_n^1-\frac{2}{2n-1}C_n^2\right\}\\&\quad+2^3\cdot(2n-3)!\left\{C_n^3-\frac{2}{2n-3}C_n^4\right\}+\cdots\end{aligned}$$

$$+2^{2t-1}\cdot(2n-2t+1)!\left\{C_n^{2t-1}\right.$$

$$\left.-\frac{2}{2n-2t+1}C_n^{2t}\right\}+\cdots \quad (5)$$

当 $1\leqslant t<n/2$ 时，

$$C_n^{2t-1}-\frac{2}{2n-2t+1}C_n^{2t}=C_n^{2t-1}\left\{1-\frac{2}{2n-2t+1}\frac{n-2t+1}{2t}\right\}$$

$$>C_n^{2t-1}\left\{1-\frac{1}{2t}\right\},$$

由以上两式立即推出，式(5)中每项均是正的，且有

$$\begin{aligned}S&>2\cdot(2n-1)!\cdot C_n^1\cdot(1/2)\\&\quad+2^3\cdot(2n-3)!\cdot C_n^3(3/4)+\cdots\\&>(1/2)\cdot(2n)!\end{aligned}$$

这就证明了式(2)．证毕．事实上，为了证明本题只要计算式(1)的前两项，是很简单的。

## （二）罗华章参赛时的解答

**第1题解答**　由于 $352=3\cdot117+1$，$1989=17\circ117$，所以{352,353,…,1989}中每个数均可唯一表为如下形式：

$$q\cdot117+r,\quad 1\leqslant r\leqslant117,\quad 3\leqslant q\leqslant16.$$

我们先把这些数按以下方式分在各个子集 $A_i$ 中：

$q\cdot117+i\in A_i$，　当 $q$ 为奇数；

$q\cdot117+(118-i)\in A_i$，　当 $q$ 为偶数。

这样，已对每个子集 $A_i$ 选定了14个元素，在每个 $A_i$ 中，

这14个数之和是

$$\sum_{k=1}^{7}[(2k+1)\cdot 117+i]$$

$$+\sum_{k=1}^{7}[(2k+2)\cdot 117+(118-i)]=16387,$$

都是相等的。下面再把剩下的数$\{1,2,\cdots,351\}$来分到各个 $A_i$ 中去。我们用这样的办法来分：先把这些数中不被 3 整除的 $2\cdot 117$个数按从小到大的顺序排列为$\{x_1,x_2,\cdots,x_{117},y_1,y_2,\cdots,y_{117}\}$；再把其中 3 的倍数按从大到小的顺序排列为$\{z_1,z_2,\cdots,z_{117}\}$；然后，把 $x_i,y_i,z_i$ 这三个数分到子集 $A_i$。我们来证明这样分成的 117 个互不相交的子集满足题目的要求。从以上的讨论知，只要证明 $x_i+y_i+z_i$ 为定值。

当 $i=1$ 时，$x_1=1,y_1=176$，$z_1=351,x_1+y_1+z_1=528$。假设当 $i=k$ 时，$x_k+y_k+z_k=528$，我们来证明当 $i=k+1$ 时也一定有 $x_{k+1}+y_{k+1}+z_{k+1}=528$。由于 $z_k$ 是 3 的倍数，所以 $x_k+y_k$ 也是 3 的倍数，因此，仅有两种情形：(i) $x_k\equiv 1 \pmod 3$，$y_k\equiv 2 \pmod 3$；(ii) $x_k\equiv 2 \pmod 3$，$y_k\equiv 1 \pmod 3$。在情形(i)将有

$$x_{k+1}=x_k+1,\quad y_{k+1}=y_k+2,\quad z_{k+1}=z_k-3,$$

而在情形(ii)将有

$$x_{k+1}=x_k+2,\quad y_{k+1}=y_k+1,\quad z_{k+1}=z_k-3.$$

因此，总有

$$x_{k+1}+y_{k+1}+z_{k+1}=x_k+y_k+z_k=528.$$

这就证明了对所有的 $1\leqslant i\leqslant 117$，总有

$$x_i + y_i + z_i = 528.$$

证毕。

**第 2 题解答** (1) 记 $\triangle ABC$ 的内心为 $I$，外心为 $O$（见图 1）.易见 $IB \perp A_0B, IC \perp A_0C$，于是 $I, B, A_0, C$ 四点共圆.因此

$$\angle BIA_0 = \angle BCA_0 = \frac{1}{2}(\pi - \angle ACB)$$

$$= \frac{1}{2}(\pi - \angle AA_1B) = \frac{1}{2}(\angle BIA_0 + \angle IBA_1).$$

由此推出

$$\angle BIA_0 = \angle IBA_1,$$

于是 $BA_1 = IA_1$。又因

$$\angle BA_0I = \frac{\pi}{2} - \angle BIA_0 = \frac{\pi}{2} - \angle IBA_1 = \angle A_0BA_1,$$

于是 $BA_1 = A_1A_0$。结合前式得到 $A_1I = A_1A_0$，因此四边形 $IBA_0C$ 的面积

$$S_{IBA_0C} = 2S_{IBA_1C}.$$

同理可证

$$S_{ICB_0A} = 2S_{ICB_1A},$$

$$S_{IAC_0B} = 2S_{IAC_1B}.$$

把上面三式相加，知(1)成立。

(2) 设 $\triangle ABC$ 的三个内角分别为 $\alpha, \beta, \gamma$。显然有 $\angle B_0OA_1 = \angle A_1OC = \alpha$。从而

$$S_{OBA_1C} = R^2 \sin\alpha, \quad S_{\triangle OBC} = \frac{1}{2}R^2 \sin 2\alpha,$$

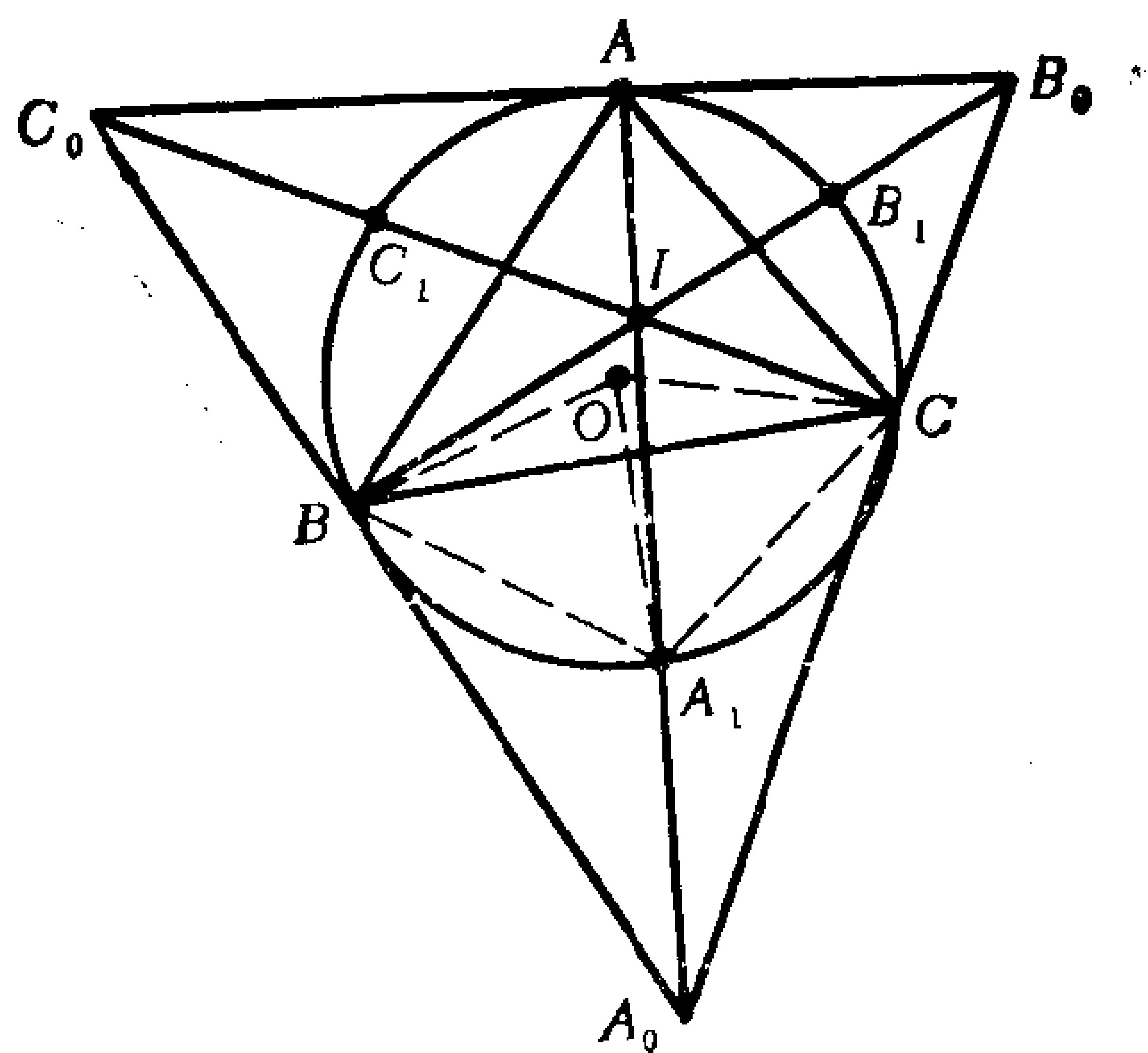



图 1

式中$R$表示$\triangle ABC$的外接圆半径。同理可证

$$S_{OAB_1C}=R^2\sin\beta,\quad S_{\triangle OAC}=\frac{1}{2}R^2\sin2\beta$$

以及

$$S_{OAC_1B}=R^2\sin\gamma,\quad S_{\triangle OAB}=\frac{1}{2}R^2\sin2\gamma.$$

由此推出

$$S_{\triangle ABC}=\frac{1}{2}R^2(\sin2\alpha+\sin2\beta+\sin2\gamma)$$

及

$$S_{AC_1BA_1CB_1}=R^2(\sin\alpha+\sin\beta+\sin\gamma).$$

据(1)由后式又得

$$S_{\triangle A_0B_0C_0}=2R^2(\sin\alpha+\sin\beta+\sin\gamma).$$

因为

$$\sin 2\alpha+\sin 2\beta=2\sin(\alpha+\beta)\cos(\alpha-\beta)$$
$$\leqslant 2\sin(\alpha+\beta)=2\sin\gamma,$$

同理又有

$$\sin 2\beta+\sin 2\gamma\leqslant 2\sin\alpha,\quad \sin 2\gamma+\sin 2\alpha\leqslant 2\sin\beta.$$

三式相加，得

$$\sin 2\alpha+\sin 2\beta+\sin 2\gamma\leqslant\sin\alpha+\sin\beta+\sin\gamma,$$

由此得

$$S_{\triangle ABC}\leqslant\frac{1}{4}S_{\triangle A_0B_0C_0},$$

(2)得证。

**第3题解答** 条件(1)是多余的。对集合$S$中一个点$P$，由条件(2)知$S$中至少有$k$个点与$P$等距，我们取定这样的一组点，设为$P_1,\cdots,P_t,t\geqslant k$。作向量$\overrightarrow{PP_1},\cdots,\overrightarrow{PP_t}$，得到了这样一个有向图$G$，$P$为向量的起点，$P_j(1\leqslant j\leqslant t)$都为终点。现在对$S$中的每个点$V_i(i=1,2,\cdots,n)$作这样的有向图。在所有这些有向图中，以$V_i$为起点的向量个数设为$a_i$，以$V_i$为终点的向量个数设为$b_i$。显见，必有等式

$$\sum_{i=1}^{n}a_i=\sum_{i=1}^{n}b_i,$$

且由条件(2)知，$a_i\geqslant k(1\leqslant i\leqslant n)$。

对每一点$V_i$以它为圆心作圆，半径为它的有向图中向量的长度。这样得到的$n$个圆的交点按重数计（即一个点是两个圆的交就算一次，当它是$k$的圆的共同交点时就算$\binom{k}{2}$次）不会超过$2\binom{n}{2}$。再注意到对每个点$V_i$，它作为两圆

的交点被计算了$\binom{b_i}{2}$次。所以必有

$$2\binom{n}{2}\geqslant\sum_{i=1}^{n}\binom{b_i}{2}.$$

由前面的讨论知

$$\sum_{i=1}^{n}\binom{b_i}{2}=\frac{1}{2}\sum_{i=1}^{n}b_i^2-\frac{1}{2}\sum_{i=1}^{n}b_i\geqslant\frac{1}{2n}\left(\sum_{i=1}^{n}b_i\right)^2-\frac{1}{2}\sum_{i=1}^{n}b_i$$

$$=\frac{1}{2n}\left(\sum_{i=1}^{n}a_i\right)^2-\frac{1}{2}\sum_{i=1}^{n}a_i\geqslant\frac{1}{2n}(nk)^2-\frac{1}{2}(nk).$$

由以上两式得 $2n-2\geqslant k^2-k$，即 $(k-1/2)^2\leqslant 2n-7/8$，这就推出所要结论。

**第4题解答** 如图2所示，以 $A,B,P$ 为圆心，分别以 $R=AD$，$r=BC$，$h$ 为半径作圆，则三圆两两外切，且$\odot P$与 $CD$ 相切。作 $AD_1\perp CD$，$BC_1\perp CD$，$PQ\perp CD$，其垂足分别为 $D_1,C_1,Q$。并记 $AD_1=R'$，$BC_1=r'$。则有

$$C_1D_1=\sqrt{(R+r)^2-(R'-r')^2},$$

$$C_1Q=\sqrt{(r+h)^2-(r'-h)^2},$$

$$D_1Q=\sqrt{(R+h)^2-(R'-h)^2}.$$

于是有

$$\sqrt{(R+r)^2-(R'-r')^2}=\sqrt{(r+h)^2-(r'-h)^2}+\sqrt{(R+h)^2-(R'-h)^2}.$$

两边平方后整理得

$$(R-h)(r-h)+(R'-h)(r'-h)-2h^2$$

$$=\sqrt{(r+h)^2-(r'-h)^2}\sqrt{(R+h)^2-(R'-h)^2}。\quad (*)$$

过$\odot A$,$\odot B$切点$K$作直线$l \parallel CD$，易见$\odot P$含于$l$与$CD$两直线之间，于是有$R+R'\geqslant 2h, r+r'\geqslant 2h$。($2h$为$\odot P$的直径。)再由$R'\leqslant R, r'\leqslant r$，得

$$(R'-h)(r'-h)\leqslant(R-h)(r-h),$$

$$(r+h)^2-(r'-h)^2\geqslant(r+h)^2-(r-h)^2=4rh,$$

$$(R+h)^2-(R'-h)^2\geqslant(R+h)^2-(R-h)^2=4Rh。$$

代入(*)式得 $2(R-h)(r-h)-2h^2\geqslant 4\sqrt{Rr}\cdot h$，配方后即为

$$Rr\geqslant(\sqrt{Rh}+\sqrt{rh})^2,$$

故

$$\frac{1}{\sqrt{h}}\geqslant\frac{1}{\sqrt{R}}+\frac{1}{\sqrt{r}}=\frac{1}{\sqrt{AD}}+\frac{1}{\sqrt{BC}}。$$

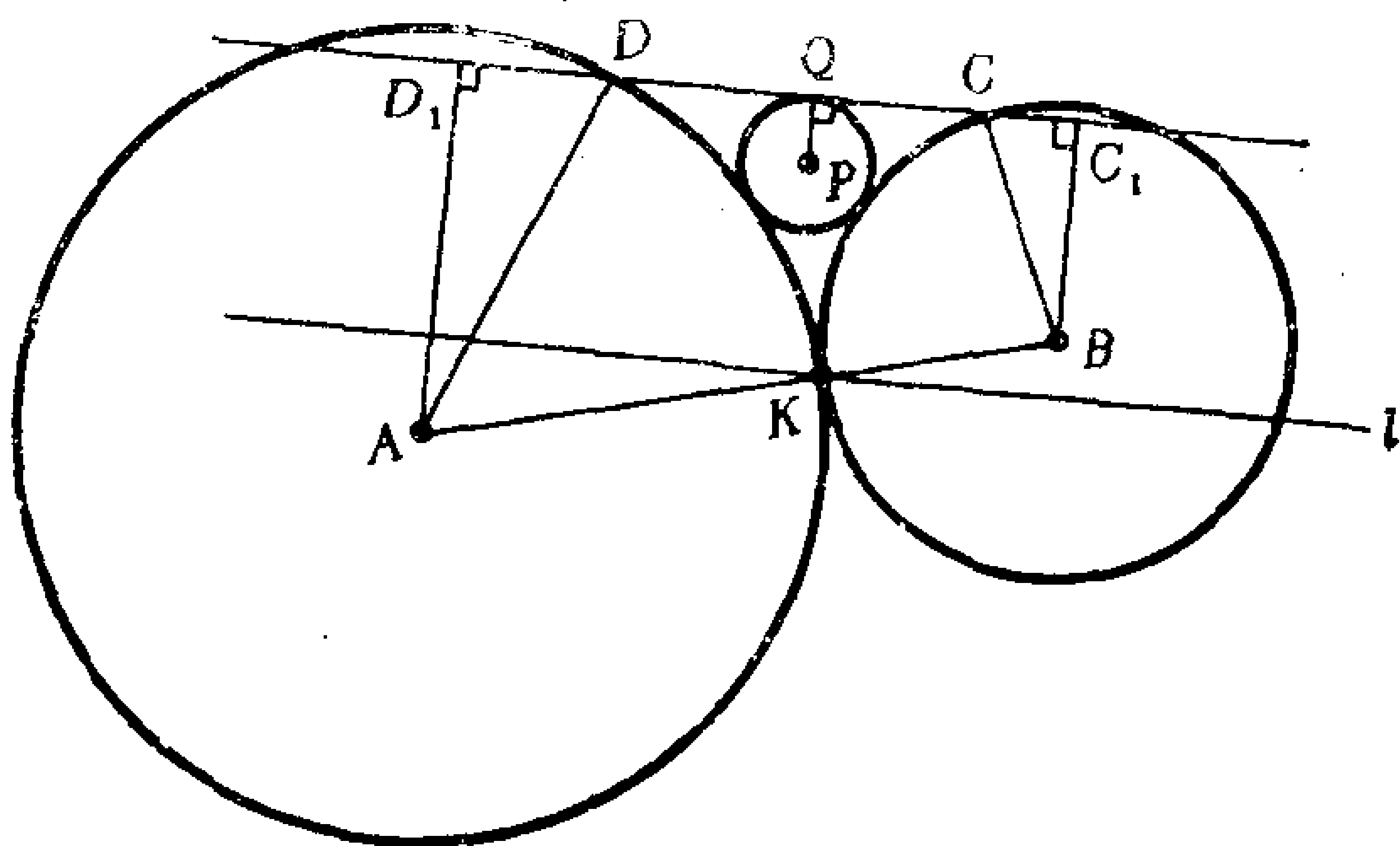



图 2

**第5题解答** 取$2n$个互不相同的素数$p_1, q_1, p_2, q_2, \cdots, p_n, q_n$。由孙子定理知同余方程组

$$x\equiv -j \pmod{p_j q_j},\quad j=1,2,\cdots,n,$$

有解，可取正整数$N$满足这同余方程组，这样，$N+1, N+2,$

$\cdots, N+n$ 这 $n$ 个相邻正整数就都不是素数的方幂，因为 $p_j q_j | N+j$。

**第6题解答** 记具有性质 $P$ 的排列个数为 $H_n$。一般的，设 $1 \leqslant k \leqslant n$，集合 $\{1,2,\cdots,k,n+1,n+2,\cdots,n+k\}$ 的具有性质 $P$ 的排列——即 $\langle x_1,\cdots,x_{2k}\rangle$ 中至少有一个 $i(1\leqslant i\leqslant 2k-1)$ 使 $|x_i - x_{i+1}| = n$ 成立的排列的个数记为 $H_k$。我们来证明：$H_k > (2k)!/2$。由此即得所要结论。当 $k=1$ 时，$H_1 = 2$ 结论显然成立。假设对 $k=j, 1\leqslant j\leqslant n-1$，结论成立，我们来推出结论对 $k=j+1$ 也成立。

首先注意到这样一个事实(*)：对任一 $x$ 属于集合 $\{1,2,\cdots,2n\}$，存在唯一的属于这集合的 $y$，使得 $|x-y|=n$。

(i) 在集合 $\{1,2,\cdots,j,n+1,\cdots,n+j\}$ 的每一个排列中，嵌入 $j+1, n+j+1$，且使这两数相邻，这样共可得到 $2(2j+1)\cdot(2j)!$ 个具有性质 $P$ 的集合 $\{1,2,\cdots,j+1,n+1,\cdots,n+j+1\}$ 的排列。(ii) 对集合 $\{1,2,\cdots,j,n+1,\cdots,n+j\}$ 的每一个具有性质 $P$ 的排列，嵌入 $j+1, n+j+1$，使得这两个数不相邻，但所得的集合 $\{1,2,\cdots,j+1,n+1,\cdots,n+j+1\}$ 的排列仍具有性质 $P$，这种嵌入至少有 $2j(2j-1)$ 种。所以用这样的方法得到的集合 $\{1,2,\cdots,j+1,n+1,\cdots,n+j+1\}$ 的具有性质 $P$ 的排列数 $\geqslant 2j(2j-1)H_j$。综合(i)，(ii)得到

$$H_{j+1} \geqslant 2j(2j-1)H_j + 2(2j+1)\cdot(2j)!.$$

由此及假设知

$$H_{j+1} > j(2j-1)\cdot(2j)! + 2(2j+1)\cdot(2j)!$$
$$> (2j+2)!/2.$$

这就证明了所要的结论。

# 第三十一届国际数学奥林匹克
# 竞 赛 试 题

**编者按** 第三十一届国际数学奥林匹克竞赛于1990年7月在北京举行。我国选手取得了非常好的成绩：获团体总分第一，五位选手获得金牌，一位获得银牌。下面是竞赛试题。我们请张筑生同志写了一份试题详解。同时，根据我国六位选手参赛时的答卷，由潘承彪同志整理编写了一份他们的有特色的解法介绍(和张筑生同志相同的就不介绍了)。在此，向为我们提供答卷的我国代表队，特别是六位选手表示感谢！

**第一题** 圆的两弦 $AB$ 与 $CD$ 交于圆内一点 $E$。设 $M$ 是弦 $AB$ 上严格在 $E$ 与 $B$ 之间的一点。过 $D,E,M$ 作一圆，设在 $E$ 点与该圆相切的直线分别与直线 $BC$ 和 $CA$ 相交于点 $F$ 和 $G$。设 $\frac{AM}{AB}=t$，试求 $\frac{EG}{EF}$(用 $t$ 来表示)。

(原题由印度提供，经选题委员会修改。)

**第二题** 设 $n\geqslant 3$。考察在圆周上给定的由 $2n-1$ 个不相同的点组成的集合 $E$。同时，考察将 $E$ 中 $k$ 个点染黑的染色办法。如果某种染色办法使得某两个染黑的点之间所夹的

弧之一的内部恰含有$E$中的$n$个点，那么我们就说这染色办法是“好的”。试求具有以下性质的最小的$k$：将$E$中任意的$k$个点染黑的染色办法都是“好的”。

（原题由捷克和斯洛伐克提供，经选题委员会修改）

**第三题** 试决定能使$\frac{2^n+1}{n^2}$是整数的一切整数$n>1$。

（罗马尼亚供题）

**第四题** 设$Q^+$是正有理数的集合。试构造一个函数$f$：$Q^+\to Q^+$，满足这样的条件：

$$f(xf(y))=\frac{f(x)}{y},\qquad \forall x,y\in Q^+. \tag{4.1}$$

（土耳其供题）

**第五题** 给定了一个初始整数$n_0>1$，两人$A$与$B$开始做游戏，他们按照以下规则轮流取数$n_1,n_2,n_3,\cdots$：

知道了$n_{2k}$，游戏者$A$选取一个整数$n_{2k+1}$，满足条件

$$n_{2k}\leqslant n_{2k+1}\leqslant n_{2k}^2.$$

知道了$n_{2k+1}$，游戏者$B$选取整数$n_{2k+2}$，使得$\frac{n_{2k+1}}{n_{2k+2}}$是一个素数的正整数次方幂。

根据约定，游戏者$A$取到数1990就获胜，游戏者$B$取到数1就获胜。

问对怎样的$n_0$，

(a) $A$有必胜策略；

(b) $B$有必胜策略；

(c) 两游戏者都无必胜策略？

（联邦德国供题）

**第六题** 证明存在具有以下性质(i)和(ii)的凸1990边形：

(i) 这多边形的各内角相等；

(ii) 这多边形各边的长度是$1^2, 2^2, \cdots, 1989^2, 1990^2$的某一排列。

(荷兰供题)

# 第三十一届IMO竞赛试题详解

张筑生（北京大学数学系）

第三十一届国际数学奥林匹克（IMO）于1990年7月8日至7月18日在北京举行。分别于7月12日和7月13日举行的两场各$4\frac{1}{2}$小时的竞赛无疑是本届IMO活动的中心。按照IMO的传统，竞赛试题事先由东道国向各参赛队征集，然后由东道国的选题委员会进行初步筛选。本届IMO共收到从35个国家（或地区）寄来的108道题。选题委员会对这些题目进行了认真细致的研究，主要做了以下三项工作：

一、改正了许多题目及原题所附解答中的错误；

二、对某些题目进行了修改，使之更适于作为IMO的竞赛题（本届IMO最后选定的六道竞赛题中，就有两道是经过选题委员会修改的题目）；

三、为大多数题目提供了更简明更富有启发性的解法，对另外一些题目补充了新解法。

选题委员会经认真研究从108道题中筛选出28道题推荐给主试委员会。IMO的主试委员会由各参赛队的领队与东道国选派的主席组成。主试委员会对本届选题委员会推荐的28道预选题进行认真热烈的讨论，然后通过表决确定六道竞赛题。下面将要介绍的就是本届IMO的六道竞赛题及这些题目的详细解答。有的竞赛题是很难的。例如本届IMO的第

六题。虽然参赛的308名选手是世界各国中学生中的佼佼者，但基本做对第六题的（该题得7分或6分者）只有19人，只占参赛人数的6％。对这样的题目，本文将尽力说明解题的思路，希望能对年轻的朋友们有所帮助。

**第一题解答** 用线段将点$D$与点$A,B$和$M$联结起来（见图1）。因为

$$\angle ECF=\angle MAD,$$
$$\angle CEF=\angle DEG=\angle EMD,$$

所以

$$\triangle CEF\backsim\triangle AMD.$$

由此得到

$$CE\cdot MD=EF\cdot AM. \tag{1.1}$$

另一方面，因为

$$\angle ECG=\angle MBD,$$
$$\angle CGE=\angle CEF-\angle GCE$$

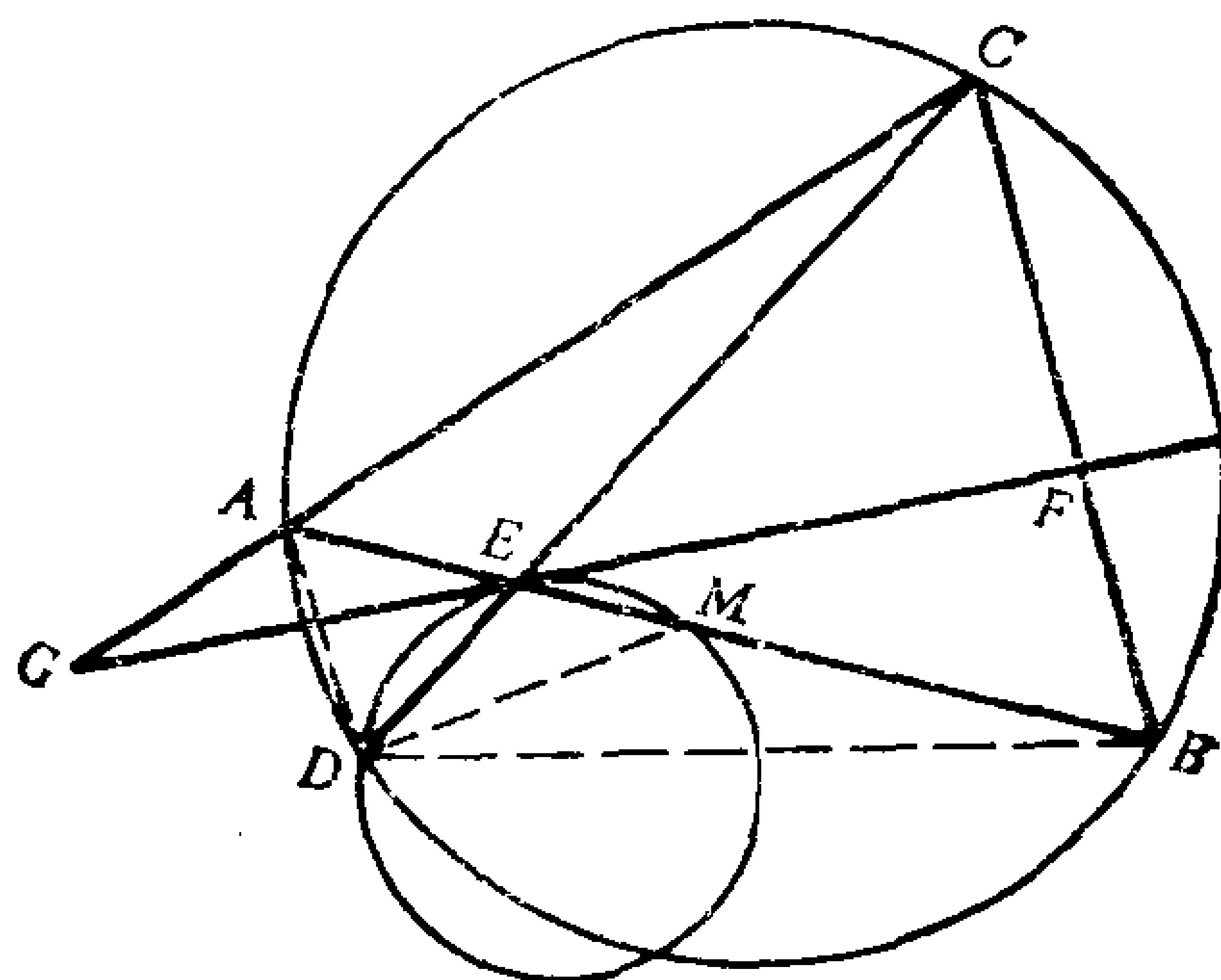



图 1

$$= \angle AMD - \angle MBD$$
$$= \angle BDM,$$

所以

$$\triangle CGE \backsim \triangle BDM,$$

因而有

$$GE \cdot MB = CE \cdot MD. \tag{1.2}$$

由(1.1)和(1.2)立即得到

$$GE \cdot MB = EF \cdot AM.$$

由此，进一步得到

$$\frac{GE}{EF} = \frac{AM}{MB} = \frac{tAB}{(1-t)AB} = \frac{t}{1-t}.$$

**注记** 在上面的讨论中，我们用到了这样的事实：点 $G$ 在 $CA$ 延长线上。这一事实可证明如下。因为$M$严格在$E$与$B$之间，所以

$$\angle BDE > \angle MDE,$$

于是有

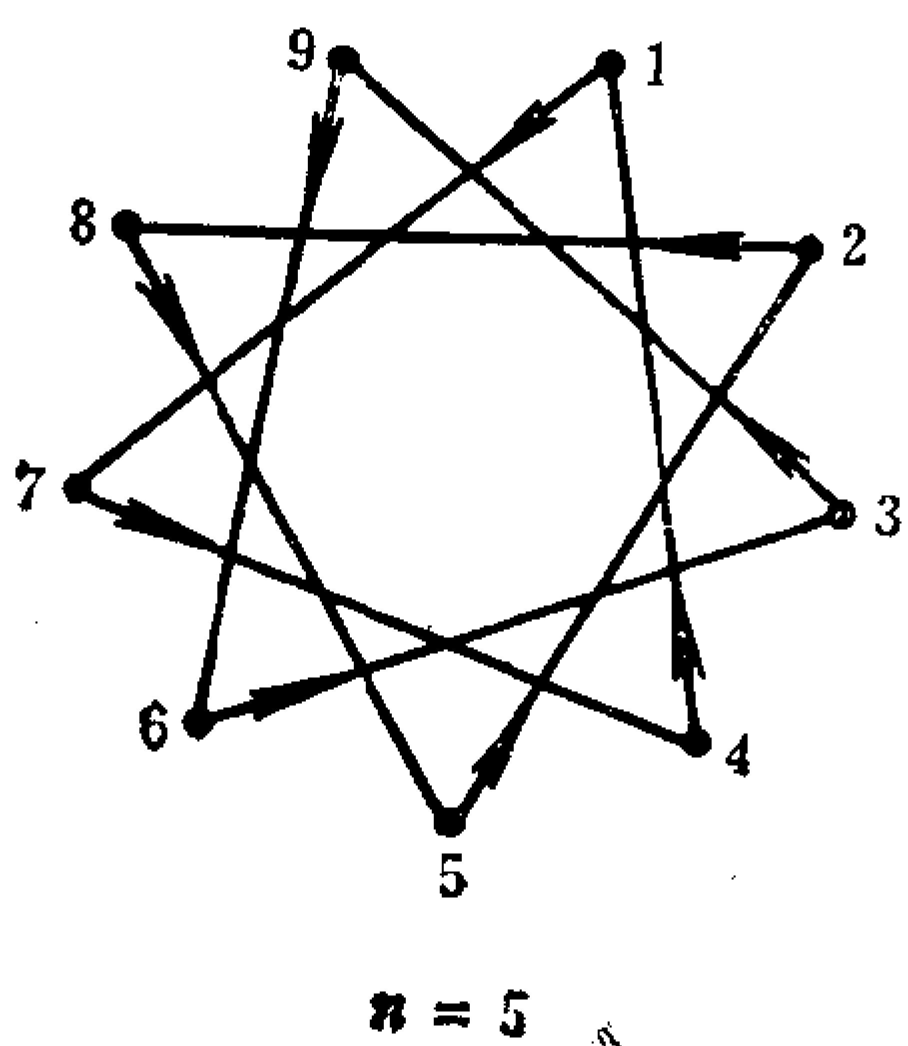
$$\angle BAC = \angle BDE > \angle MDE = \angle BEF.$$

因而直线 $CA$ 与 $FE$ 的交点 $G$ 应该在线段 $CA$ 的延长线上。

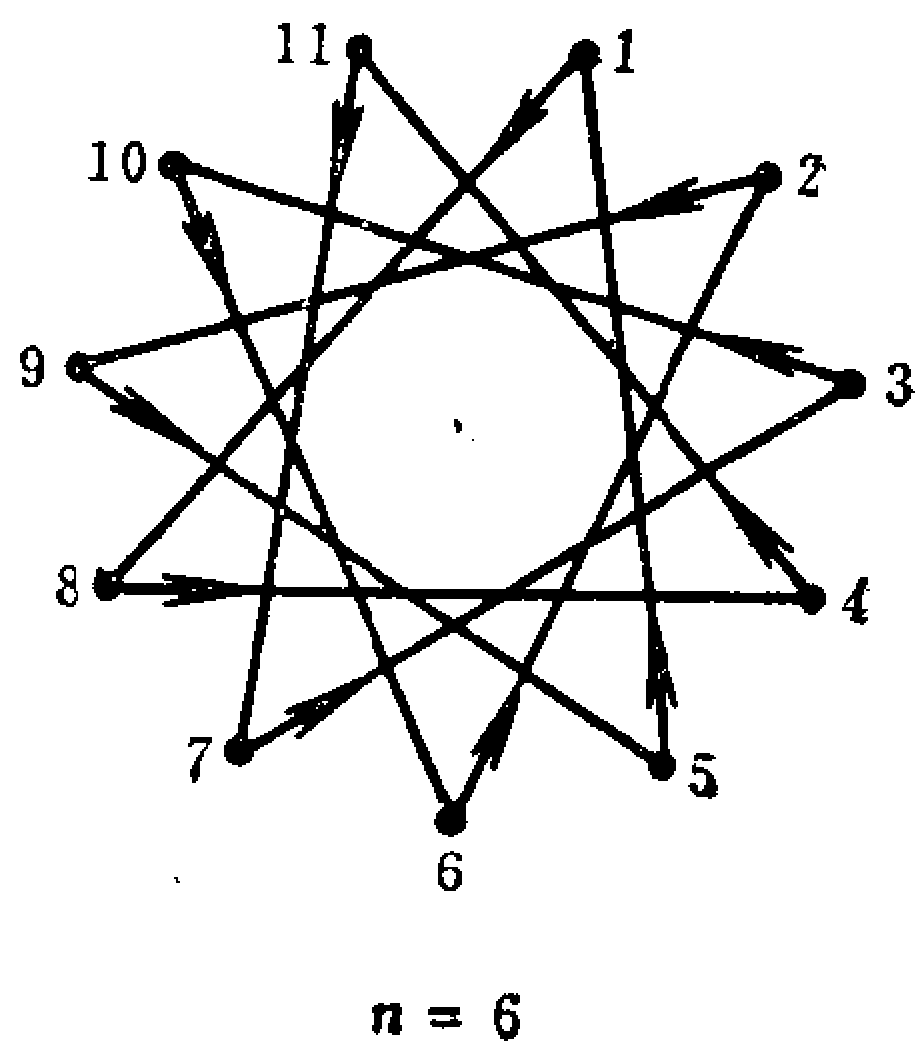
**第二题解答** $E$中的两个点称为是"相关的"，如果这两点所夹的弧段之一的内部恰含有$E$的$n$个点。沿顺时针方向依此用$1,2,\cdots,2n-1$给$E$的点标号（从任意指定一点开始）。这样，与点$i$相关的点仅有两个：$i+n+1$（当$i+m+1>2n-1$时是$i+n+1-(2n-1)$），及$i+n-2$（当$i+n-2>2n-1$时是$i+n-2-(2n-1)$）。所以，$E$中两点相关的充要条件是它们的标号相差$n+1$或$n-2$，本题就是要决定

具有以下性质的自然数 $k$ 的最小值：$E$ 的任意 $k$ 个点中至少有两点是相关的。

现将 $E$ 中任意两个相关的点都用线段联结起来，我们就得到以 $E$ 中的点为顶点，以这些线段为边的图 $G$。由于每一点有且仅有两个点和它相关，所以图 $G$ 在其每一个顶点 $p$ 处的度数 $d(p)$（即以 $p$ 为一个端点的边的个数）都等于2。从任一顶点 $i$ 出发，沿图 $G$ 的边联结与 $i$ 相关的点 $i+n+1$（标号见前约定），依次这样联结，那么若干步后必回到原顶点 $i$，这是因为总共只有有限个点，且每点的度数均为2。这样就得到了图 $G$ 的一个子图，我们称它为"圈"。由于顶点度数均为2，所以任意两个不同的圈（如果有的话）一定没有公共顶点。因而，图 $G$ 是由一些（可以是一个）没有公共顶点的圈组成，相关的点就是同一个圈上相邻的点（即一条边的两端点）。附图画出了：$n=5$ 时，图 $G$ 由三个圈：$\{1,7,4,1\}$；$\{2,8,5,2\}$，$\{3,9,6,3\}$ 组成；$n=6$ 时，图 $G$ 只有一个圈：$\{1,8,4,11,7,3,10,6,2,9,5,1\}$。



$n=5$



$n=6$

下面来证明：对任意两个点 $1 \leqslant i, j \leqslant 2n-1$，它们在同一个圈上的充要条件是存在整数 $x, y$，使得

$$i-j=x(n+1)+y(2n-1)。$$

条件的必要性由圈的构造法及标号的约定立即推出。下证充分性。不妨设 $x \leqslant 0$（为什么？）。若 $x=0$，则 $y=0$，即 $i=j$ 为同一点，结论当然成立。若 $x<0$。因为 $i$ 和 $i_1=i+(n+1)$ 或 $i'_1=i_1-(2n-1)=i+(n+1)-(2n-1)$（当 $i+(n+1)>2n-1$ 时）在同一个圈上，所以有

$$i_1-j=(x+1)(n+1)+y(2n-1)$$

或

$$i'_1-j=(x+1)(n+1)+(y-1)(2n-1)。$$

依此下去，总可得到一点 $i'$，$1 \leqslant i' \leqslant 2n-1$，它和 $i, i_1, i_2 \cdots$ 在同一个圈上，且满足

$$\begin{aligned} i'-j &= (x+(-x))(n+1)+y'(2n-1) \\ &= y'(2n-1), \end{aligned}$$

这表明 $i'$ 和 $j$ 为同一点。这就证明了充分性。

由于

$$\begin{aligned} (n+1, 2n-1) &= (n+1, 2n-1-2(n+1)) \\ &= (n+1, -3) = (n+1-3n, -3) \\ &= (3, 2n+1), \end{aligned}$$

以下分两种情形来讨论：

**情形 I** $3 \nmid 2n-1$。这时 $(n+1, 2n-1)=1$，所以必有整数 $s, t$ 使得

$$s(n+1)+t(2n-1)=1。$$

因此，对任意两点 $1 \leqslant i, j \leqslant 2n-1$，必有整数 $x, y$，使得

$$i-j=x(n+1)+y(2n-1).$$

所以，$E$中所有点在同一个圈上，即$G$是由单独一个圈构成。沿这圈每隔一个顶点取一个顶点，可得$\left[\frac{2n-1}{2}\right]=n-1$个两两不相邻——即不相关的点。但是，任取$n$个点，就必定会出现一对相邻——即相关的点。所以，这时$k$的最小值等于$n$。

情形Ⅱ　$3|2n-1$。这时$(n+1,2n-1)=3$。所以一定存在整数$s,t$使

$$s(n+1)+t(2n-1)=3.$$

因此，对两个点$1\leqslant i,\ j\leqslant 2n-1$存在$x,y$，使得有

$$i-j=x(n+1)+y(2n-1)$$

成立的充要条件是$3|i-j$。这样就推出：1,2,3这三个点属于不同的圈，以及任一点必属于其中的一个圈。因而，$G$由三个两两没有公共顶点的圈组成，每个圈含有$(2n-1)/3$个顶点。在每个圈上分别每隔一个顶点取一个顶点，可在每个圈上得到$\left[\frac{(2n-1)/3}{2}\right]$个两两不相邻——即不相关的点，这样，总共得到

$$3\left[\frac{(2n-1)/3}{2}\right]=3\cdot\frac{(2n-1)/3-1}{2}=n-2$$

个两两不相关的点。但是，任取$n-1$个点，就必定会有$\left[\frac{(2n-1)/3}{2}\right]+1$个点属于同一个圈（为什么?），因此，必有两个点相邻——即相关。所以，这时$k$的最小值等于$n-1$。

**注记**　本题若利用同余概念来做，证明可以表述得更简洁。

**第三题解答**　首先，因为$2^n+1$是奇数，所以满足要求的$n$必须是奇数。其次，显然$n=3$是本题的一个解。下面将证明本题无其他解，因而$n=3$是唯一解。因为任何奇整数$n$都可表示成

$$n=3^l\cdot m \quad (l\geqslant 0,\ 2\nmid m,\ 3\nmid m),$$

所以只须证明：对于能使得$n^2|2^n+1$成立的$n=3^l\cdot m$，必有

（Ⅰ）$l\leqslant 1$，

（Ⅱ）$m=1$。

为了证明（Ⅰ），将利用以下事实：

**引理1**　对于奇整数$s>0$，必有

$$2^{2s}-2^s+1\equiv 3 \pmod 9.$$

**引理1的证明**　这引理的结论可以通过观察以下表格而得到：

$2^{2t}-2^t+1$的mod 9剩余

| $t$ | $2^{2t}$ | $-2^t$ | 1 | $2^{2t}-2^t+1$ |
|---|---|---|---|---|
| 1 | 4 | 7 | 1 | 3 |
| 2 | 7 | 5 | 1 | |
| 3 | 1 | 1 | 1 | 3 |
| 4 | 4 | 2 | 1 | |
| 5 | 7 | 4 | 1 | 3 |
| 6 | 1 | 8 | 1 | |
| 7 | 4 | 7 | 1 | 3 |

——这表格是循环的，从$t=7$起重复出现从$t=1$开始的

情形，这就证明了引理1。

**（I）之证明** 设 $l \geqslant 1$，则有

$$
\begin{aligned}
2^n+1 &= 2^{3^l \cdot m}+1 \\
&= (2^{3^{l-1}\cdot m}+1)(2^{2\cdot 3^{l-1}\cdot m}-2^{3^{l-1}\cdot m}+1) \\
&= \cdots\cdots\cdots\cdots \\
&= 3\cdot\prod_{k=0}^{l-1}(2^{2\cdot 3^k\cdot m}-2^{3^k\cdot m}+1).
\end{aligned}
$$

因为 $n=3^l \cdot m$，$3^{2l} \mid n^2$，所以

$$3^{2l} \mid 2^n+1=2^{3^l\cdot m}+1.$$

由引理1可知

$$3^l \Bigm| \prod_{k=0}^{l-1}(2^{2\cdot 3^k\cdot m}-2^{3^k\cdot m}+1),$$

$$3^{l+1} \nmid \prod_{k=0}^{l-1}(2^{2\cdot 3^k m}-2^{3^k\cdot m}+1).$$

因而有

$$3^l \mid 3.$$

由此得知 $l \leqslant 1$。

为了证明（II），将利用以下简单事实：

**引理2** 对于给定的自然数 $a$ 和 $h$，设 $j$ 是最小的自然数，使得

$$a^j \equiv -1 \pmod h.$$

如果非负整数 $r<j$ 使得

$$a^r \equiv \pm 1 \pmod h,$$

那么必有 $r=0$。

**引理2的证明** 对于

$$a^r \equiv -1 \pmod{h}$$

的情形，利用 $j$ 的最小性就能得到结论 $r=0$，对于

$$a^r \equiv 1 \pmod{h}$$

的情形，我们仍利用 $j$ 的最小性。注意到

$$a^{j-r} \equiv a^{j-r} \cdot a^r \equiv a^j$$
$$\equiv -1 \pmod{h}$$

以及

$$0 \leqslant r < j,$$

即可断定 $r=0$。

（Ⅱ）之证明 （用反证法） 若 $m \neq 1$，设 $p \geqslant 5$ 是 $m$ 的最小素因数，则

$$2^n \equiv -1 \pmod{p}.$$

设 $j$ 是最小的自然数，使得

$$2^j \equiv -1 \pmod{p}.$$

如果

$$n = qj + r, \quad 0 \leqslant r < j,$$

那么

$$(-1)^q 2^r \equiv 2^{qj} \cdot 2^r \equiv -1 \pmod{p},$$

即

$$2^r \equiv (-1)^{q+1} \pmod{p}.$$

根据引理2，应有 $r=0$，由此得知

$$j \mid n.$$

另一方面，Fermat 小定理告诉我们

$$2^{p-1} \equiv 1 \pmod{p}.$$

因为 $p-1 \neq 0$，根据引理2必有

$$p-1\geqslant j.$$

但 $p\geqslant 5$ 是 $m$ 的最小素因数，而 $j|n$，所以只能有

$$j=1 \text{ 或者 } j=3.$$

对这样的 $j$，由

$$2^j\equiv -1 \pmod p$$

可得

$$p|3 \text{ 或者 } p|9.$$

但这与 $p\geqslant 5$ 矛盾。这证明了 $m=1$。

最后的结论是：唯一满足要求的解是

$$n=3.$$

**第四题解答** 首先，我们指出，所求的 $f$ 应该满足以下各条件：

(1) $f$ 是单映射。事实上，设

$$f(y_1)=f(y_2).$$

这等式两边乘以正有理数 $x$ 并以 $f$ 作用之，就可得到

$$f(xf(y_1))=f(xf(y_2)).$$

利用(4.1)又可得到

$$\frac{f(x)}{y_1}=\frac{f(x)}{y_2}.$$

由此可知

$$y_1=y_2.$$

这证明了 $f$ 是单映射。

(2) $f(1)=1$。事实上，我们有

$$f(f(1))=f(1\cdot f(1))=\frac{f(1)}{1},$$

也就是

$$f(f(1))=f(1).$$

因为 $f$ 是单映射，所以

$$f(1)=1.$$

(3) $f(f(y))=\frac{1}{y}$。这是因为

$$f(f(y))=f(1\cdot f(y))=\frac{f(1)}{y}=\frac{1}{y}.$$

(4) $f\left(\frac{1}{y}\right)=\frac{1}{f(y)}$. 这是因为

$$f\left(\frac{1}{y}\right)=f(f(f(y)))=\frac{1}{f(y)}.$$

(5) $f(x\cdot t)=f(x)\cdot f(t)$. 事实上，在(4.1)中取

$$y=f\left(\frac{1}{t}\right)$$

就得到

$$f\left(x\cdot f\left(f\left(\frac{1}{t}\right)\right)\right)=\frac{f(x)}{f\left(\frac{1}{t}\right)}=\frac{f(x)}{\frac{1}{f(t)}}$$

$$=f(x)\cdot f(t).$$

再利用(3)就得到

$$f(x\cdot t)=f(x)\cdot f(t).$$

反过来，如果函数 $f:Q^+\to Q^+$ 具有上面列举的性质(3)和(5)，那么显然有

$$f(xf(y))=f(x)\cdot f(f(y))=\frac{f(x)}{y},$$

即函数 $f$ 满足(4.1)。我们只须构造一个具有上面性质(1)—(5)的函数 $f$。由性质(4)，(5)知只须对任意素数 $p$ 规定 $f(p)$ 的值。

将全体素数按从小到大的顺序排列：

$$p_1, p_2, \cdots, p_k, p_{k+1}, \cdots.$$

首先，我们规定

$$f(p_j)=\begin{cases} p_{j+1}, & \text{如果 } j \text{ 是奇数;} \\ \dfrac{1}{p_{j-1}}, & \text{如果 } j \text{ 是偶数.} \end{cases}$$

然后，按照关系

$$f(x\cdot t)=f(x)\cdot f(t),$$

$$f\left(\frac{1}{y}\right)=\frac{1}{f(y)},$$

将 $f$ 的定义范围扩充到全体正有理数的集合 $Q^+$。显然这样定义的函数

$$f:Q^+\to Q^+$$

满足条件(3)和(5)，因而满足条件(4.1)。

**第五题解答** 我们约定记

$$W=\left\{n_0\in \boldsymbol{N}\,\middle|\,\begin{array}{l}\text{从这 } n_0 \text{ 开始，}\\ A\text{ 有必胜策略}\end{array}\right\}.$$

因为 $45^2=2025>1990$，所以显然有

$$\{45,46,\cdots,1990\}\subseteq W.$$

我们找一个自然数，要求抹去其任何一个素数方幂因子之后所得之数不小于 45。易见

$$2^2\times 3\times 5\times 7=420$$

就是一个合乎要求的数．若 $n_0\leqslant 420\leqslant n_0^2$，则 $A$ 第一次只须取 $n_1=420$。其后不论 $B$ 怎样选取，总有 $n_2\geqslant 45$。于是 $A$ 第二次选取就能获胜。因为

$$21<420<21^2,$$

所以

$$\{21,22,\cdots,44\}\subseteq W.$$

接着，我们试找一个自然数，要求抹去其任何一个素数方幂因子之后所得的数不小于 21。易见

$$2^3\times 3\times 7=168$$

合乎要求。因为

$$13<168<13^2,$$

所以

$$\{13,14,\cdots,20\}\subseteq W.$$

再找一个数，要求抹去其任何一个素数方幂因子之后所得的数不小于13。易见

$$3\times 5\times 7=105$$

合乎要求．因为

$$11<105<11^2,$$

所以

$$\{11,12\}\subseteq W.$$

再找一个数，要求抹去其任何一个素数方幂因子之后所得的数不小于11。易见

$$2^2\times3\times5=60$$

合乎要求．因为

$$8<60<8^2,$$

所以

$$\{8,9,10\}\subseteq W.$$

对于$n_0>1990$的情形，总可以选取 不 小 于 3 的自然数$m$，使得

$$2^m\times3^2<n_0\leqslant2^{m+1}\times3^2<n_0^2.$$

于是，$A$可取$n_1=2^{m+1}\times3^2$。这时不论$B$怎样选择，总有

$$8\leqslant n_2<n_0.$$

若$n_2>1990$，则可用$n_2$代替 $n_0$。游戏者$A$重复上面所说的策略，又能使得

$$8\leqslant n_4<n_2.$$

在有限步之后，$A$总能使$n_{2k}$下降到这样一种情形：

$$8\leqslant n_{2k}\leqslant1990.$$

由此得知：$n_0>1990$时，$A$总可获胜。

若$n_0\leqslant5$，而$n_1$满足$n_0\leqslant n_1\leqslant n_0^2$，则$n_1$至多只有两个不同的素数因子(因为最小三个相异素数之积 $2\times3\times5=30>5^2$)，并且只要$n_1$不是完全平方数就一定有 $n_1$的一个素数方幂因子

$$p^r > \sqrt{n_1}.$$

于是，$B$可取

$$n_2 = \frac{n_1}{p^r} < \sqrt{n_1} \leqslant n_0.$$

继续这样的策略，$B$能使$n_{2k}$逐步递降，直到取得1．因此，只要$n_0 \leqslant 5$，游戏者$B$总可获胜．注意这里所取$n_1$至多只有两个不同的素因子是关键．

对于$n_0 = 6$或者$n_0 = 7$的情形，$n_0^2 = 36$或者$n_0^2 = 49$．$A$应选取介于$n_0$与$n_0^2$之间的至少有三个相异的素数因子之数（根据刚才的分析，选取少于三个素数因子之数将导致$A$的失败）．于是，$n_1$的合理选择只能是

$$n_1 = 2 \times 3 \times 5 \quad 或者 \quad n_1 = 2 \times 3 \times 7.$$

$B$取$n_2$的正确策略应是取$n_1$抹去最大素数因子后得到的数（否则将有$n_2 \geqslant 8$，对$B$不利）．于是，$B$只能选取$n_2 = 6$．接下去双方不得不轮流选取

$$30, 6, 30, 6, 30, 6, \cdots.$$

游戏将不分胜负．

综上所述，我们得出结论：

(a) $n_0 \geqslant 8$时$A$有必胜策略，

(b) $n_0 \leqslant 5$时$B$有必胜策略，

(c) $n_0 = 6$或7时，双方均无必胜策略．

**第六题解答** 我们首先通过分析将问题转化成更易于处理的形式．假设满足上述条件的1990边形存在．沿反时针方向给这多边形的各边定向，再将各边的起点移到原点，这样得到1990个向量．相邻的两边对应的两向量相邻，它们之间

的夹角为$\alpha=\frac{2\pi}{1990}$(这是因为凸多边形的每个内角均为$\pi-\alpha$)。以复数表示平面向量，原问题转化为：求证存在具有以下性质(1)，(2)和(3)的1990个复数。

(1) 相邻两复数之间的夹角为$\alpha$；

(2) 各复数的长度是$1^2,2^2,\cdots,1990^2$的某一排列；

(3) 这些复数之和等于0。

这就是说，我们需要求得$1^2,2^2,\cdots,1990^2$的一个排列$n_0,n_1,\cdots,n_{1989}$，使得

$$\sum_{s=0}^{1989} n_s e^{is\alpha}=0.$$

如果将这些复数的长度$n_s$看成“重量”，那么问题又可转述为：给定了一个水平放置的单位圆，设法将$1^2,2^2,\cdots,1990^2$这些“重量”按某种次序放到等分圆周的1990个点上，要求这系统的重心落到圆心上。下面，我们就来解决这一问题。

第一步。依次将$1^2,2^2,\cdots,1990^2$这些“重量”每两个分成一组，这样得到995组：

$$\{1^2,2^2\},\ \{3^2,4^2\},\cdots,\{1989^2,1990^2\}.$$

将同一组中的两个“重量”放到单位圆周的某一对对径点上。至于哪一组放到哪一条直径的两端，则待下面的讨论来确定。这样，各组中两复数之和的长度分别为

$$3,7,11,\cdots,3979$$

(首项为3，公差为4的等差数列)。

于是，问题进一步转化为：将3，7，11，…，3979这些“重量”，放到等分圆周的995个点上，要求重心落到圆心上。

第二步．我们注意到

$$995=5\times199.$$

由此得到启发，再将3,7,11,…,3979这些“重量”每五个分成一组，共分成199组：

$$(*)\qquad\begin{cases}\{3,7,11,15,19\},\ \{23,27,31,35,39\},\\ \{43,47,51,55,59\},\cdots,\\ \qquad\cdots,\{3963,3967,3971,3975,3979\}.\end{cases}$$

记$\beta=\frac{2\pi}{199},\gamma=\frac{2\pi}{5}$。我们把顶点在

$$1,e^{i\gamma},e^{2i\gamma},e^{3i\gamma},e^{4i\gamma}$$

的正五边形记为$F_1$，并把正五边形$e^{ik\beta}F_1$记为$F_{k+1}$。依次将(*)中所列的199组“重量”放到$F_1,F_2,\cdots,F_{199}$这些正五边形的顶点上，我们得到分成199组的995个复数，其中第$k+1$组为

$$(20k+3)e^{ik\beta},\qquad (20k+7)e^{i(k\beta+\gamma)},$$
$$(20k+11)e^{i(k\beta+2\gamma)},\qquad (20k+15)e^{i(k\beta+3\gamma)},$$
$$(20k+19)e^{i(k\beta+4\gamma)}.$$

五次单位根$e^{i\gamma}$有这样的性质：

$$1+e^{i\gamma}+e^{2i\gamma}+e^{3i\gamma}+e^{4i\gamma}=0.$$

因而第$k+1$组中的五个复数之和可以化简为

$$\eta e^{ik\beta},$$

这里

$$\eta=3+7e^{i\gamma}+11e^{2i\gamma}+15e^{3i\gamma}+19e^{4i\gamma}.$$

于是，所有这199组共995个复数之总和为

$$\eta(1+e^{\beta i}+\cdots+e^{198\beta i})=0.$$

我们证明了：存在满足条件(1)，(2)和(3)的1990个复数。因而，确实存在满足题目条件(i)和(ii)的凸1990边形。

最后，我们指出，可以将上面的解答简单地整理成以下几行式子再加上几句说明的话。

$$0=\sum_{k=0}^{198}\sum_{l=0}^{4}(20k+4l+3)e^{i(k\beta+l\gamma)}$$

$$=\sum_{k=0}^{198}\sum_{l=0}^{4}[(10k+2l+2)^2-(10k+2l+1)^2]e^{i(k\beta+l\gamma)}$$

$$=\sum_{k=0}^{198}\sum_{l=0}^{4}\sum_{m=1}^{2}(10k+2l+m)^2e^{i(k\beta+l\gamma+m\pi)};$$

当指标$k$遍历$0,1,\cdots,198$，指标$l$遍历$0,1,\cdots,4$，指标$m$遍历1，2的时候，表示式

$$10k+2l+m$$

遍历从1到1990的所有自然数。与此同时，表示式

$$e^{i(k\beta+l\gamma+m\pi)}$$

$$=e^{i\frac{10k+398l+995m}{1990}2\pi}$$

遍历$1,e^{\alpha i},\cdots,e^{1989\alpha i}$这1990个复数。

**注记** 最后，我们来说明所作的1990边形确实是凸的。

为此，只须指出这样的事实：无论将这封闭折线的哪一条边延长成直线，折线其余所有的顶点都位于该直线的同一侧。

设所作的封闭折线是$A_0A_1\cdots A_{1989}A_0$。在上面的讨论中，

我们已将向量$\overrightarrow{A_sA_{s+1}}$表示为复数：

$$\overrightarrow{A_sA_{s+1}}=n_s e^{is\alpha},$$

其中

$$\alpha=\frac{2\pi}{1990},\qquad A_{1990}=A_0.$$

必要时可以将这图形 $A_0A_1\cdots A_{1989}A_0$ 旋转 $\alpha$ 的适当倍数并相应地改变各顶点的编号，总可以将我们所关心的任何一边重新标号为$A'_0A'_1$，并且仍可设

$$\overrightarrow{A'_sA'_{s+1}}=n'_s e^{is\alpha}.$$

以下为简便起见，我们省去新记号中的撇“′”，直截了当地把这封闭折线的任意一边当作$A_0A_1$。并且，还可以认为$A_0=O$是坐标原点，$\overrightarrow{A_0A_1}=\overrightarrow{OA_1}$指向实轴的正方向。我们还约定，允许用同一记号来记平面上的点和该点所代表的复数。

于是，我们可以写

$$A_k=\overrightarrow{OA_k}=\sum_{r=0}^{k-1}\overrightarrow{A_rA_{r+1}}$$

$$=\sum_{r=0}^{k-1}n_r e^{ir\alpha}=-\sum_{s=k}^{1989}n_s e^{is\alpha}.$$

对于 $0<r<k\leqslant 995$，显然有

$$\sin r\alpha=\sin\frac{r}{1990}2\pi>0,$$

因而

$$\operatorname{Im}(A_k)=\sum_{r=0}^{k-1}n_r\operatorname{Im}(e^{ir\alpha})$$

$$=\sum_{r=0}^{k-1} n_r \sin r\alpha$$

$$>n_1 \sin \alpha>0.$$

对于 $996 \leqslant k \leqslant s \leqslant 1989$，显然有

$$-\sin s\alpha=-\sin \frac{s}{1990} 2\pi>0,$$

因而

$$\begin{aligned} \operatorname{Im}(A_k) &= -\sum_{s=k}^{1989} n_s \operatorname{Im}(e^{is\alpha}) \\ &= -\sum_{s=k}^{1989} n_s \sin s\alpha \\ &> -n_{1989} \sin 1989\alpha \\ &= n_{1989} \sin \alpha>0. \end{aligned}$$

我们看到，折线其余所有的顶点都位于 $A_0A_1$ 所在直线的同一侧，而 $A_0A_1$ 可以是这封闭折线的任何一条边。因此，所作的1990边封闭折线确实构成一个凸多边形。

# 第三十一届IMO我国选手解答介绍

潘承彪　整理

我国参赛的六位选手是：汪建华、周彤、库超、王崧、余嘉联及张朝晖。除了张筑生同志提供的解答之外，他们在参赛时还给出了不同的解法。下面是他们关于第一、二、三、六题的一些有特色的解法介绍。在编写时作了适当整理，主要介绍解题的思想和关键步骤，一些具体推导将留给读者。

## 第　一　题

**余嘉联、汪建华的解法**（图同张筑生的解法）

（1）由$M$点严格在$EB$内知$\angle MEF<\angle MAC$，所以点$G,A$在点$C$的同侧。同理，点$F,B$在点$C$同侧。

（2）由面积公式、正弦定理等推出

$$\frac{GE}{EF}=\frac{\sin\angle ECG}{\sin\angle ECF}\cdot\frac{\sin\angle EFC}{\sin\angle EGC}.$$

同理推出：

$$\frac{BM}{AM}=\frac{\sin\angle BDM}{\sin\angle ADM}\cdot\frac{\sin\angle BAD}{\sin\angle ABD}.$$

（3）由关于弦切角、圆周角的定理推出

$$\angle BDM = \angle CGE, \quad \angle ADM = \angle CFE.$$

(4)

$$\frac{GE}{EF} = \frac{AM}{BM} = \frac{t}{1-t}.$$

**库超的解法**

(1) 作 $AP$ 平行于 $GF$，$AP$ 分别交 $CD$、$CB$ 于点 $H$，$P$ (见图1)，得出

$$\frac{GE}{EF} = \frac{AH}{PH}.$$

图 1

(2) $\angle BEF = \angle EDM$，$\angle HAE = \angle MDE$。由此推出 $A$，$H$，$M$，$D$ 四点共圆。

(3) 由(2)及 $A$，$B$，$C$，$D$ 四点共圆推出 $\angle EHM = \angle EAD = \angle ECB$，因而 $HM$ 和 $BC$ 平行，并有

$$\frac{AM}{BM} = \frac{AH}{PH}.$$

(4)

$$\frac{GE}{EF}=\frac{AM}{BM}=\frac{t}{1-t}.$$

## 第 二 题

我国选手本题的解法基本上相同，都用了同余的概念和性质。特别是汪建华、王崧的解法论证严格，表述简洁。下面介绍他们的解法。

（1）按顺时针方向，从某点开始依次将这 $2n-1$ 个点编号为 $1,2,\cdots,2n-1$。并继续往下重复编号，$2n,2n+1,\cdots$。这样，每个点有无穷多个编号，但只要

$$i\equiv j\pmod{2n-1},$$

$i$ 和 $j$ 就代表同一个点，反过来也对。

（2）当 $3\nmid 2n-1$ 时，$(2n-1,n-2)=1$。在这情形，我们从点 1 开始，依此将点 $a_r=1+r(n-2)$ 与点 $a_{r+1}=1+(r+1)(n-2)$ 连线，取 $r=0,1,\cdots,2n-2$。由于 $(2n-1,n-2)=1$，所以，当 $r$ 遍历模 $2n-1$ 的一个完全剩余系 $0,1,\cdots,2n-2$ 时，$1+r(n-2)$ 也遍模 $2n-1$ 的完全剩余系。因此，从点 1 开始的这一条连线联结了所有这 $2n-1$ 个点，且每点恰出现一次。注意到 $a_{2n-1}$ 和 $a_0$ 是同一点，所以，这一连线形成一个回路 $T$。

（3）将点 $a_0,a_1,\cdots,a_{2n-2}$ 中任意 $n$ 个点染黑，则必有两个相邻点 $a_{r_0},a_{r_0+1}$ 同时被染黑，因此，这种染色方式一定是好的。

（4）将 $a_1,a_3,\cdots,a_{2n-3}$ 这 $n-1$ 个点染黑，这种染色方

式不是好的。

(5) 当$3\nmid 2n-1$时，$k$的最小值为$n$。

(6) 当$3|2n-1$时，$(2n-1,n-2)=3$。记点

$$a_{1,r}=1+r(n-2),$$
$$a_{2,r}=2+r(n-2),$$
$$a_{3,r}=3+r(n-2).$$

(7) 将点$a_{1,r}$与点$a_{1,r+1}$相连，当依次取$r=0,1,\cdots,\frac{(2n-1)}{3}-1$时得到一个回路$T_1$，所有模3余1的点在$T_1$中恰出现一次。同样的，将点$a_{2,r}$与$a_{2,r+1}$相连，得到回路$T_2$，所有模3余2的点在$T_2$中恰出现一次；将点$a_{3,r},a_{3,r+1}$相连，得到回路$T_3$，所有模3余0的点在$T_3$中恰出现一次。此外，回路$T_1,T_2,T_3$两两没有公共点，每个回路中有$(2n-1)/3$个点。

(8) 若将点$a_0,a_1,\cdots,a_{2n-2}$中的任意$n-1$个点染黑，则必有一个回路$T_j$中有$\left[\frac{n-1}{3}\right]+1=\frac{n-2}{3}+1$个点；进而推出在这回路$T_j$中必有相邻两点$a_{j,r_0},a_{j,r_0+1}$都被染黑。因此，这种染色方式一定是好的。

(9) 若将$a_{j,1},a_{j,3},\cdots,a_{j,(2n-7)/3}$，$j=1,2,3$，这$n-2$个点染黑，则这种染色方法不是好的。

(10) 当$3|2n-1$时，$k$的最小值为$n-1$。

## 第 三 题

**本题汪建华、王崧、余嘉联、张朝晖、及库超的解法基本相同。我们适当整理，介绍如下：**

（1）若 $n$ 是解，则必是奇数，所以可表为

$$n=3^l m,\quad l\geqslant 0, 3\nmid m, 2\nmid m. \qquad (*)$$

为方便起见，我们来求 $n\geqslant 1$ 的全部整数解①。

（2）用归纳法证明：对任意整数 $k\geqslant 1$ 必有

$$3^k \mid 2^{3^{k-1}}+1,\qquad 3^{k+1}\nmid 2^{3^{k-1}}+1.$$

（3）设 $k,s$ 是正整数。那么，$3^k\mid 2^s+1$ 的充要条件是

$$2\nmid s,\qquad 3^{k-1}\mid s.$$

（4）若 $n$ 是解，则在表示式（$*$）中的 $l\leqslant 1$。

（5）设 $q>1$ 是奇数，且存在正整数 $d$ 使得

$$q\mid 2^d+1.$$

再设 $d_0$ 是使上式成立的最小正整数。那么，若对正整数 $c$ 有

$$q\mid 2^c+1 \text{ 或 } q\mid 2^c-1,$$

则必有 $d_0\mid c$（参看张筑生的解法）。

（6）对任意奇素数 $p$，必有 $p\mid 2^{p-1}-1$。

（7）若 $n$ 是解，则表示式（$*$）中的 $m=1$。用反证法。若 $m>1$，则 $m$ 的素因子必大于 3。设 $p$ 是 $m$ 的最小素因子，$d_0$ 是使 $p\mid 2^d+1$ 的最小正整数。证明必有

$$d_0\mid n,\qquad d_0\mid p-1.$$

利用（4）推出：必有 $d_0=1$ 或 3。但这和 $p$ 是大于 3 的素数矛盾。

---

① 这不是他们的提法。这里只是为了以下叙述方便。

**周彤的解法**

他在证明中用了 Euler 定理：设 $m \geqslant 1$，$(a,m)=1$，则有 $a^{\phi(m)} \equiv 1 (\bmod m)$，这里 $\phi(m)$ 是 Euler 函数，表示 1,2,…,$m$ 中和 $m$ 互素的数的个数。下面介绍周彤解法的思想，但不用 Euler 定理。

(1) 对任意奇整数 $m$，必有 $1 \leqslant d \leqslant m$ 使得

$$2^d \equiv 1 (\bmod m).$$

设 $d_0$ 是使上式成立的最小正整数。那么，如果

$$2^l \equiv 1 (\bmod m),$$

则必有 $d_0 | l$。

(2) 设 $m>1$，则 $2^m \not\equiv 1 (\bmod m)$。用反证法。若不然，则有 $m$ 是奇数及

$$2^m \equiv 1 (\bmod m). \qquad (**)$$

设 $p$ 是 $m$ 的最小素因子。$d_0$ 是使

$$2^{d_0} \equiv 1 (\bmod p)$$

成立的最小正整数。由于 $p>1$ 及 $2^{p-1} \equiv 1 (\bmod p)$，所以

$$1 < d_0 \leqslant p-1 < p.$$

因此，$d_0$ 必有素因子 $p_1 < p$。但另一方面，$2^m \equiv 1 (\bmod p)$，由(1)知 $d_0 | m$，$p_1 | m$。这和 $p$ 的最小性矛盾。

**注记** 周彤是用 Euler 定理证(2)的。证明如下：设 $m_1$ 是式(**)成立的大于 1 的最小正整数。再设 $d_1$ 是使

$$2^{d_1} \equiv 1 \ (\bmod m_1)$$

成立的最小正整数。由(1)知 $d_1 | m_1$，因而有

$$2^{d_1}\equiv 1\ (\text{mod}\ d_1).$$

由 $m_1$ 的最小性知 $d_1=1$ 或 $m_1$。由 $m_1>1$ 知 $d_1\neq 1$。由$d_1$ 的最小性及 Euler定理知$d_1\leqslant\phi(m_1)$。当$m_1>1$时有 $m_1>\phi(m_1)$，所以 $d_1\neq m_1$。因此式(* *)不能成立。

(3) 由(1),(2)推出，当$m$是大于1的奇数时，(1)中的 $d_0$ 满足$1<d_0<m$。

(4) 设正整数 $n_0$ 满足

$$2^{2n_0}\equiv 1\ \ (\text{mod}\ n_0), \qquad (* *)'$$

$t_0$ 是使

$$2^{t_0}\equiv 1\ \ (\text{mod}\ n_0)$$

成立的最小正整数．若 $n_0>1$，则 $t_0$ 必为偶数。因为，由(1)知 $t_0|2n_0$。若 $t_0$ 为奇数，则 $t_0|n_0$，因而有

$$2^{n_0}\equiv 1\ \ (\text{mod}\ n_0),$$

但由(2)知这是不可能的。设 $t_0=2n_1$，则有 $n_1|n_0$。此外，由(3)知 $t_0<n_0$，所以 $n_1<n_0$。显见，$n_1$ 亦满足式(* *)′，即

$$2^{2n_1}\equiv 1\ \ (\text{mod}\ n_1).$$

(5) 由(4)的讨论知，只要 $n_1>1$，就可重复上面的步骤，得到 $n_2|n_1$，$n_2<n_1$，且满足

$$2^{2n_2}\equiv 1\ (\text{mod}\ n_2).$$

这样，经有限步后就得到了一串正整数 $n_0>n_1>\cdots>n_k>n_{k+1}=1$，满足 $n_{j+1}|n_j$ 及

$$2^{2n_{j+1}}\equiv 1\ \ (\text{mod}\ n_j), \qquad j=0,1,\cdots,k.$$

由 $n_{k+1}=1$ 推出 $n_k=3$，进而有 $n_{k-1}|2^6-1=63$。

（6）设 $n>1$ 是本题的解．显见有

$$2^n\equiv-1(\mathrm{mod}\ n),\qquad 2^{2n}\equiv1(\mathrm{mod}\ n).$$

取(5)中的 $n_0=n$。若 $n_0>3$，则由(5)知 $n_{k-1}|n_0$。下面来讨论 $n_{k-1}$ 的可能的取值。

由$n_k=3|n_{k-1}$及$n_{k-1}|63$知：(a)若$7|n_{k-1}$,则$21|n_{k-1}$，$21|n$。因而有

$$2^n\equiv(2^3)^{n/3}\equiv1(\mathrm{mod}\ 7),$$

这和由 $2^n\equiv-1(\mathrm{mod}\ n)$ 推出的 $2^n\equiv-1(\mathrm{mod}\ 7)$矛盾．(b)若$7\nmid n_{k-1}$，则必有$n_{k-1}=9$。因此$n_0$必可表为

$$n=n_0=3^lm,\quad l\geqslant2,\ 2\nmid m,\ 3\nmid m.$$

下面来证这是不可能的。

（7）设 $l$ 是正整数．那么使

$$2^h\equiv1(\mathrm{mod}\ 3^l)$$

成立的最小的 $h$ 为$2\cdot3^{l-1}$．而且有

$$2^{2\cdot3^{l-1}}\not\equiv1(\mathrm{mod}\ 3^{l+1}).$$

用归纳法证。

（8）若$n=n_0>3$是解，则有

$$2^{2n}\equiv1(\mathrm{mod}\ n^2).$$

由(6)知必有$9|n$。而由(1),(7)知上式对这样的 $n$ 不可能成立。所以，仅有解$n=3$。

## 第　六　题

本题的解法都是归结为求1,2,…,1990 的一个排列：$r_1$,

$r_2,\cdots,r_{1990}$，使得

$$\sum_{j=1}^{1990} r_j^2 e^{i\alpha j}=0,$$

这里$\alpha=2\pi/1990$。注意到$e^{2\pi i}=1$，更一般的可归为求1,2,…,1990的一个排列$r_1,r_2,\cdots,r_{1990}$，及模1990的一个完全剩余系$u_1,u_2,\cdots,u_{1990}$，使得

$$\sum_{j=1}^{1990} r_j^2 e^{i\alpha u_j}=0。\qquad (*)$$

下面介绍汪建华、周彤和王崧的解法，王崧的解法与众不同。

**汪建华的解法** 设

$$a_{k,j}=10k+j,\ 1\leqslant j\leqslant 10,\ 0\leqslant k\leqslant 198.$$

$$b_{h,l}=10h+199l,\ 0\leqslant l\leqslant 9,\ 0\leqslant h\leqslant 198.$$

显见，$a_{k,j}$恰好取1,2,…,1990这些值，且每个一次。由初等数论知，$b_{h,l}$恰好是模1990的一个完全剩余系。

考虑

$$\begin{aligned}T_k=&a_{k,1}^2e^{i\alpha b_{k,0}}+a_{k,2}^2e^{i\alpha b_{k,2}}+a_{k,3}^2e^{i\alpha b_{k,4}}\\&+a_{k,4}^2e^{i\alpha b_{k,6}}+a_{k,5}^2e^{i\alpha b_{k,8}}+a_{k,6}^2e^{i\alpha b_{k,5}}\\&+a_{k,7}^2e^{i\alpha b_{k,7}}+a_{k,8}^2e^{i\alpha b_{k,9}}+a_{k,9}^2e^{i\alpha b_{k,1}}\\&+a_{k,10}^2e^{i\alpha b_{k,3}}.\end{aligned}$$

显见，

$$T=\sum_{k=0}^{198}T_k$$

就给出了式(*)左边形式的和式。下面来证明$T=0$。容易算出

$$T_k = -5e^{i10k\alpha}\sum_{j=1}^{5}(2j-1)e^{i2\pi(j-1)/5},$$

右边的和式是与$k$无关的常数。由此即得$T=0$。

**周形的解法** 设

$$a_{j,k}=199j+k,\quad 0\leqslant j\leqslant 9,\ 1\leqslant k\leqslant 199.$$

显见，$a_{j,k}$恰好每个一次地取1,2,…,1990这些值。再设

$$b_{0,k}=a_{0,k},\quad b_{1,k}=a_{6,k},\quad b_{2,k}=a_{2,k},\quad b_{3,k}=a_{8,k},$$
$$b_{4,k}=a_{4,k},\quad b_{5,k}=a_{5,k},\quad b_{6,k}=a_{1,k},\quad b_{7,k}=a_{7,k},$$
$$b_{8,k}=a_{3,k},\quad b_{9,k}=a_{9,k}.$$

并约定

$$b_{j,k}=b_{l,k},\quad l\equiv j\pmod{10}.$$

对每一个$k$取定一个整数$m_k$，考虑和式

$$S_k=\sum_{j=0}^{9}b_{j+m_k,k}^2\,e^{i\alpha a_{j,k}}.$$

显见

$$S=\sum_{k=1}^{199}S_k$$

就给出了形如式(*)左边的和式。容易算出

$$S_k=\left\{5\cdot 199^2\sum_{j=1}^{4}2j\,e^{i\alpha 199j}\right\}e^{i\alpha(199m_k+k)}.$$

{…}内是和$k$无关的常数。下面来确定$m_k$的取值。由于当$0\leqslant m\leqslant 9$，$1\leqslant k\leqslant 199$时，$199m+k$恰好每个一次地取1,2,…,1990这些值。因此，一定可以取到$0\leqslant m_k\leqslant 9$，使得

$c_k=199m_k+k$，当 $k=1,2,3,\cdots,199$时，恰好取$10,20,30,\cdots,1990$这199个数(次序可以不同)。由此即得 $S=0$。

**注记** 请读者比较汪建华、周彤的解法的异同。

**王崧的解法** 考虑多项式

$$G(x)=x^{4\cdot 199}+x^{3\cdot 199}+x^{2\cdot 199}+x^{199}+1,$$

$$H(x)=x^{5\cdot 198}+x^{5\cdot 197}+\cdots+x^{5\cdot 2}+x^{5}+1,$$

以及

$$M_1(x)=G(x)(199x^{198}+198x^{197}+\cdots+2x+1),$$

$$M_2(x)=H(x)(4\cdot 199x^3+3\cdot 199x^2+2\cdot 199x+199),$$

$$M(x)=M_1(x)+M_2(x).$$

容易看出：$M(x)$是994次多项式，系数必在$1,2,\cdots,995$这些值之中。我们来证明$M(x)$的系数两两不同，因此恰好$1,2,\cdots,995$各出现一次。

(1) $M_1(x)$中出现$x^j$项，$j=199l+k$，$0\leqslant l\leqslant 4$，$0\leqslant k\leqslant 198$，且它的系数$a_j=k+1$。所以在$M_1(x)$中$x^j(0\leqslant j\leqslant 994)$都出现，且系数$a_j$满足$1\leqslant a_j\leqslant 199$。

(2) $M_2(x)$中出现$x^j$项，$j=5k+l$，$0\leqslant l\leqslant 3$，$0\leqslant k\leqslant 198$，且它的系数$b_j=199(l+1)$。

(3) $M(x)$中没有两项的系数相同。用反证法。假若$x^{j_1}$和$x^{j_2}$的系数$c_{j_1}$和$c_{j_2}$相等，$j_1\neq j_2$。我们有

$$c_{j_1}=a_{j_1}+b_{j_1},\qquad c_{j_2}=a_{j_2}+b_{j_2}.$$

因此，

$$a_{j_2}-a_{j_1}=b_{j_1}-b_{j_2}.$$

由(1)知，必有

$$0\leqslant|a_{j_2}-a_{j_1}|\leqslant 198.$$

若$a_{j_2}-a_{j_1}=0$，则由(1)推出$199|(j_2-j_1)$。另一方面，这时必有$b_{j_2}=b_{j_1}$，由(2)知，$5|(j_2-j_1)$。因此推出$995|(j_2-j_1)$。但已知$0\leqslant j_1,j_2\leqslant 994$，矛盾。

进而考虑多项式

$$N(x)=4M(x)-(x^{994}+x^{993}+\cdots+x+1)。$$

$N(x)$是994次多项式，$x^j$的系数是$4c_j-1$，$0\leqslant j\leqslant 994$，$c_j$恰好每个一次地取值$1,2,\cdots,995$。因此

$$N(x)=\sum_{j=0}^{994}(4c_j-1)x^j$$

$$=\sum_{j=0}^{994}(2c_j)^2x^j-\sum_{j=0}^{994}(2c_j-1)^2x^j。$$

取$x=e^{2i\alpha}$，

$$N(e^{2i\alpha})=\sum_{j=0}^{994}(2c_j)^2e^{i\alpha(2j)}+\sum_{j=0}^{994}(2c_j-1)^2e^{i\alpha(2j+995)}$$

显见，右边就是形如(*)的和式。但另一方面，由$N(x)$的定义容易推出$N(e^{2i\alpha})=0$。

# 初等数学问题(1)解答

1. 如果第一组数中有三个 0，另一个不是 0，那么第二组开始的各组将全是 0. 它们与第一组不同，这时命题成立.

如果第一组数中有两个 0 。当这两个 0 不相邻时，同样出现第二组开始各组全是 0 的现象。当这两个 0 相邻时，第二组一定是 3 个 0 ，另一个不是 0 。这样，从第三组开始的各组全都是 0 。可见，第一组中有两个是 0 ，另外两个不是 0 时，命题成立。

如果第一组数中有一个 0 ，另外三个都不是 0 。那么第二组数就有两个 0（另外两个不是 0），第三组数就有三个 0（另一个数不是 0），同理可知，第一组数中有一个 0 时，命题成立。

综上，第一组数中有 0（但不全为 0）时，命题得证。

以下设第一组数中都不是 0 。考虑各组数的符号。

如果第一组数中有一个是负数，另外三个是正数，那么从第一组开始的符号将是

| | | | | |
|---|---|---|---|---|
| 第一组 | + + + − | + + − + | + − + + | − + + + |
| 第二组 | + + − − | + − − + | − − + + | − + + − |
| 第三组 | + − + − | − + − + | + − + − | − + − + |
| 第四组 | − − − − | − − − − | − − − − | − − − − |
| 第五组 | + + + + | + + + + | + + + + | + + + + |

这表明第二、三、四、五组与第一组不同，且从第五组开始的各组总是正数，也与第一组不同。因此，第一组中有一个是负数，三个是正数时，命题得证。

如果第一组数中两个负数两个正数。则是上面表中第二行或第三行的情形，同理可知命题成立。同理知，当第一组中四个都是负数时，命题也成立。

如果第一组数中有三个负数，一个正数。那么第二组数就一定是两个负数两个正数。这种情况下实际上也得到了证明。

综上，第一组数全不为0且有负数时，命题得证。

以下设第一组中四个数都是正数，且不都是1．

设第一组数为$a,b,c,d$。第二组数就是$ab,bc,cd,da$。假如

$$ab=a,\quad bc=b,\quad cd=c,\quad da=d,$$

那么就有$a=b=c=d=1$，与题设矛盾。这表明第二组数与第一组数不同。

考查各组数的积形成的数列：

$$abcd,(abcd)^2,(abcd)^4,\cdots,(abcd)^{2^{n-1}},\cdots$$

如果有$k(k\geqslant3)$，使得第$k$组数与第一组数相同，那么一定有

$$abcd=(abcd)^{2^{k-1}},$$

所以

$$abcd=1.$$

等式$abcd=(ab)\cdot(cd)=(ad)\cdot(bc)=1$表明：在$ab,cd$中一定有一个不少于1，另一个不多于1。不妨设$ab\geqslant1$，$cd\leqslant1$。

同理，在 $ad,bc$ 中也有一个不少于1，另一个不多于1。

当 $ad\geqslant1,bc\leqslant1$ 时，为书写方便起见，记 $ab=\alpha,ad=\beta$，则 $cd=\frac{1}{\alpha}$，$bc=\frac{1}{\beta}$。

如果 $\alpha\geqslant\beta\geqslant1$，则 $1\geqslant\frac{1}{\beta}\geqslant\frac{1}{\alpha}$。从第二组数列写下去，各组数列是

$$\alpha,\frac{1}{\beta},\frac{1}{\alpha},\beta; \qquad \text{（第二组）}$$

$$\frac{\alpha}{\beta},\frac{1}{\alpha\beta},\frac{\beta}{\alpha},\alpha\beta; \qquad \text{（第三组）}$$

$$\frac{1}{\beta^2},\frac{1}{\alpha^2},\beta^2,\alpha^2; \qquad \text{（第四组）}$$

$$\frac{1}{\alpha^2\beta^2},\frac{\beta^2}{\alpha^2},\alpha^2\beta^2,\frac{\alpha^2}{\beta^2}; \qquad \text{（第五组）}$$

$$\frac{1}{\alpha^4},\beta^4,\alpha^4,\frac{1}{\beta^4}; \qquad \text{（第六组）}$$

………………

容易看出从第二组开始各组中的最大数依次是

$$\alpha,\alpha\beta,\alpha^2,\alpha^2\beta^2,\alpha^4,\cdots$$

显然

$$\alpha\leqslant\alpha\beta\leqslant\alpha^2\leqslant\alpha^2\beta^2\leqslant\alpha^4\leqslant\cdots.$$

这表明，当 $\alpha\geqslant\beta\geqslant1$ 时，从第二组数开始的各组数中的最大数是（非严格）递增的。

如果 $\beta\geqslant\alpha\geqslant1$ 时，各组最大数是

$$\beta, \alpha\beta, \beta^2, \alpha^2\beta^2, \beta^4, \cdots$$

又有

$$\beta \leqslant \alpha\beta \leqslant \beta^2 \leqslant \alpha^2\beta^2 \leqslant \beta^4 \leqslant \cdots,$$

同样有这种"最大数递增"现象。根据以上分析知：当 $\beta = ad \geqslant 1, bc = \frac{1}{\beta} \leqslant 1$时，从第三组数开始，各组的最大数形成的数列是递增数列。如果 $bc \geqslant 1, ad \leqslant 1$，可设 $\beta = bc$，仍然有这种最大数递增现象。这种递增现象表明：如果某一组与第一组相同，它们的最大数也一定相同，不妨设

$$\alpha = \alpha\beta = \alpha^2 = \alpha^2\beta^2 = \alpha^4 = \cdots,$$

可见 $\alpha = \beta = 1$。所以 $\frac{1}{\alpha} = \frac{1}{\beta} = 1$，就是说第二组数都是1。由此得到 $ab = bc = cd = da = 1$，即 $a = b = c = d = 1$，与题设矛盾。这矛盾表明，不存在 $k \in N$，$k \geqslant 3$，使得第 $k$ 组数与第一组数相同。

我们已证了第二组数与第一组数不同。可见，从第二组开始的任何一组数与第一组数都不同。

2． 把 $m, m+1, m+2, \cdots, m+n$ 分解因数，集中 2 的因子。即设

$$m + i = 2^{\alpha_i} \cdot a_i \quad (i = 0, 1, \cdots, n)。$$

其中 $2 \nmid a_i, a_i \in N, \alpha_i$是非负整数。

设 $\alpha = \max\{\alpha_0, \alpha_1, \cdots, \alpha_n\}$，假设有 $i \neq j$，使 $2^\alpha \cdot a_i = 2^\alpha \cdot a_j$，不妨设 $i < j$，那么

$$m + i = 2^\alpha \cdot a_i < m + j = 2^\alpha \cdot a_j.$$

这表明 $a_i < a_j$。

但 $a_i, a_j$ 均为奇数，可见存在偶数 $t$，使得 $a_i<t<a_j$（$t$ 为偶数），那么就有

$$m+i=2^{\alpha}\cdot a_i<2^{\alpha}\cdot t<2^{\alpha}\cdot a_j=m+j.$$

这表明 $2^{\alpha}\cdot t\in\{m,m+1,\cdots,m+n\}$。但 $t$ 为偶数，所以 $2^{\alpha}\cdot t=2^{\alpha+1}\cdot\left(\frac{t}{2}\right)\left(\frac{t}{2}\in N\right)$。这表明 $\alpha<\max\{\alpha_0,\alpha_1,\cdots,\alpha_n\}$，与题设矛盾。这矛盾表明不存在 $i\neq j$，使得 $2^{\alpha}\cdot a_i=2^{\alpha}\cdot a_j$。即是说，在集合 $\{m,m+1,\cdots,m+n\}$ 各元素中，有 $2^{\alpha}$ 因数的元素仅有一个。设 $\alpha=\alpha_k\ (k\in\{0,1,\cdots,n\})$，即 $m+k=2^{\alpha}\cdot a_k$。

考虑乘积

$$(2^{\alpha-1}\cdot a_0\cdot a_1\cdot\cdots\cdot a_n)\cdot\left(\frac{1}{m}+\frac{1}{m+1}+\cdots+\frac{1}{m+k}+\cdots+\frac{1}{m+n}\right)$$

$$\begin{aligned}=&2^{\alpha-1-\alpha_0}\cdot a_1a_2\cdot\cdots\cdot a_n+2^{\alpha-1-\alpha_1}a_0\cdot a_2\cdot a_3\cdot\cdots\cdot a_n+\cdots\\&+2^{\alpha-1-\alpha_{k-1}}a_0\cdot a_1\cdot\cdots\cdot a_{k-2}\cdot a_k\cdot\cdots\cdot a_n\\&+2^{\alpha-1-\alpha_k}a_0\cdot a_1\cdot\cdots\cdot a_{k-1}\cdot a_{k+1}\cdot\cdots\cdot a_n\\&+2^{\alpha-1-\alpha_{k+1}}\cdot a_0\cdot a_1\cdot\cdots\cdot a_k\cdot a_{k+2}\cdot\cdots\cdot a_n+\cdots\\&+2^{\alpha-1-\alpha_n}\cdot a_0\cdot a_1\cdot\cdots\cdot a_{n-1},\end{aligned}$$

其中 $(\alpha-1)-\alpha_j\geqslant 0$（$j\neq k$，$j=0,1,\cdots,n$），且 $\alpha-1-\alpha_j\in Z$，$(\alpha-1)-\alpha_k=(\alpha-1)-\alpha=-1$。可见，等号右边除去

$$a^{\alpha-1-\alpha_k}\cdot a_0\cdot a_1\cdot\cdots\cdot a_{k-1}\cdot a_{k+1}\cdot\cdots\cdot a_n$$

$$=\frac{1}{a}\cdot a_0\cdot a_1\cdot\cdots\cdot a_{k-1}\cdot a_{k+1}\cdot\cdots\cdot a_n$$

这一项之外，各项都是整数。而这一项的分子是奇自然数，分母是偶数 2，因此唯独这一项不是整数。这表明乘积

$$(2^{a-1}\cdot a_0\cdot a_1\cdot\cdots\cdot a_n)\cdot\left(\frac{1}{m}+\frac{1}{m+1}+\cdots+\frac{1}{m+n}\right)$$

不是整数。但 $2^{a-1}\cdot a_0\cdot a_1\cdot\cdots\cdot a_n\in \boldsymbol{N}$，因此

$$\frac{1}{m}+\frac{1}{m+1}+\cdots+\frac{1}{m+n}\notin \boldsymbol{N}.$$

3　如果 3-$L$ 型板按(a)，(b) 两种方向使用(见图 1)，我们对 6×6 的方格盘的每格染上三色中的一种颜色（用1,2,3

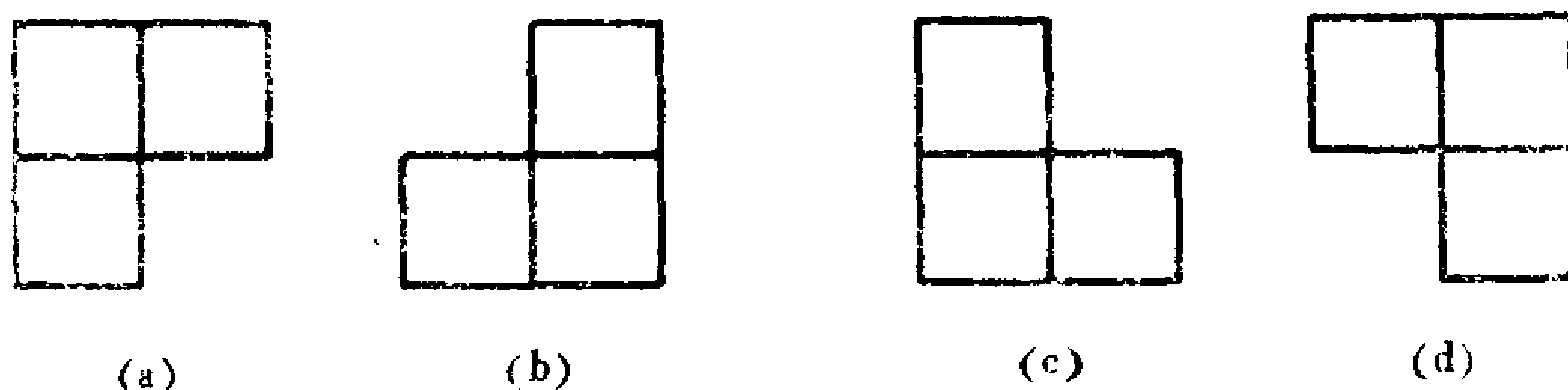



图　1

表示 3 种颜色）（见图2,图3）。如果 3-$L$ 型板按(c)，(d)方式使用，则按图 2 涂色。

| | | | | | |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 1 | 2 | 3 |
| 2 | 3 | 1 | 2 | 3 | 1 |
| 3 | 1 | 2 | 3 | 1 | 2 |
| 1 | 2 | 3 | 1 | 2 | 3 |
| 2 | 3 | 1 | 2 | 3 | 1 |
| 3 | 1 | 2 | 3 | 1 | 2 |

图　2

| | | | | | |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 1 | 2 | 3 |
| 3 | 1 | 2 | 3 | 1 | 2 |
| 2 | 3 | 1 | 2 | 3 | 1 |
| 1 | 2 | 3 | 1 | 2 | 3 |
| 3 | 1 | 2 | 3 | 1 | 2 |
| 2 | 3 | 1 | 2 | 3 | 1 |

图　3

图 2 表明：任何(a)，(b) 方向的 3-$L$ 型板盖住的三格中恰有两色，图 2 表明：任何(c)，(d) 方向的 3-$L$ 型板盖住的三格中也恰有两色。但在任何一种情况下，$3\times1$型板总是盖住三种颜色的三个小方格，因而总有不被盖住的方格。这表明：$6\times6$的方格盘不可能被一个 3-$L$ 型板和十一个$3\times1$型板完全覆盖。

4． 先证一个辅助命题：

对于任意自然数 $k$，只要 $n\geqslant k$，都有 $f(n)\geqslant k$。用数学归纳法。

(1) 当 $k=1$ 时，$\because f(n)\in \boldsymbol{N}$，$\therefore f(n)\geqslant1$，这表明 $k=1$ 时命题成立。

(2) 设有自然数 $k$，当 $n\geqslant k$ 时，$f(n)\geqslant k$。考虑 $k+1$。

当 $n\geqslant k+1$ 时，$n-1\geqslant k$。由归纳假设知，对一切 $n-1\geqslant k$，都有 $f(n-1)\geqslant k$。对于自然数 $f(n-1)$，再次由归纳假设知

$$f[f(n-1)]\geqslant k.$$

因此，

$$f(n)>f[f(n-1)]\geqslant k,$$

所以

$$f(n)\geqslant k+1.$$

这表明 $n=k+1$ 时命题也成立。

由(1)，(2)所证知：辅助命题为真。

由 $k$ 的任意性，取 $k=n$，则有 $f(n)\geqslant n$ 对一切 $n\in \boldsymbol{N}$ 成立。

另一方面，取 $n=f(k)$，不等式 $f(n)\geqslant n$ 就成为

$$f[f(k)]\geqslant f(k),$$

可见

$$f(k+1)>f[f(k)]\geqslant f(k).$$

这表明函数 $f(k)$是严格增函数。但

$$f(n+1)>f[f(n)],$$

所以

$$n+1>f(n),$$

即
$$n\geqslant f(n).$$

根据我们证出的关系

$$\begin{cases} n\geqslant f(n), \\ n\leqslant f(n) \end{cases}$$

知：对一切 $n\in N$ 都有 $f(n)=n$。

5. 设集合

$I$ ={全部选手},

$A$ ={第一题对的考生},

$B$ ={第二题对的考生},

$C$ ={第三题对的考生},

$D$ ={第四题对的考生}.

设 $n(P)$ 表示集 $P$ 中的元素数。则 $n(A)=235$，$n(A\cap B)=59$，$n(A\cap C)=29$，$n(A\cap D)=15$，$n(A\cap B\cap C\cap D)=3$。

因为

$$n(A\cap B\cap C)>n(A\cap B\cap C\cap D)=3,$$

同理，$n(A\cap C\cap D)$，$n(A\cap B\cap D)>3$，可见：

$$n(A\cap B\cap C)+n(A\cap B\cap D)+n(A\cap C\cap D)$$
$$-n(A\cap B\cap C\cap D)>6.$$

注意到

$$
\begin{aligned}
n(A\cap\overline{B}\cap\overline{C}\cap\overline{D}) &= n[A\cap(\overline{B+C+D})]\\
&= n(A+B+C+D)-n(B+C+D)\\
&= [n(A)+n(B)+n(C)+n(D)-n(A\cap B)\\
&\quad -n(A\cap C)-n(A\cap D)-n(B\cap C)-\cdots\\
&\quad +n(A\cap B\cap C)+\cdots-n(A\cap B\cap C\cap D)]\\
&\quad -[n(B)+n(C)+n(D)-n(B\cap C)-n(C\cap D)\\
&\quad -n(B\cap D)+n(B\cap C\cap D)]\\
&= n(A)-n(A\cap B)-n(A\cap C)-n(A\cap D)\\
&\quad +n(A\cap B\cap C)+n(A\cap B\cap D)+n(A\cap C\cap D)\\
&\quad -n(A\cap B\cap C\cap D)\\
&> 235-59-29-15+6=138.
\end{aligned}
$$

可见

$$n(A\cap\overline{B}\cap\overline{C}\cap\overline{D})\geqslant 139=3\times 46+1>3\times 46.$$

由抽屉原则知，存在一个国家，该国派出的选手中至少有 4 人做对了且只做对了第一题.

（周沛耕提供，陈剑刚校）

# 初等数学问题①（2）

1．若 $p$ 是一个不等于5的奇素数，

求证：$p^2-1$ 与 $p^2+1$ 中至少有一个能被10整除。

2．已知：$n$ 是不小于3的奇数。

求证：$2^{n-1}(2^n-1)\equiv 1 \pmod 9$。

3．已知数列的前14项是：4，6，9，10，14，15，21，22，25，26，33，34，35，38，…。

求它的第15，16，17项。

4．已知 $n$ 是一个素数，$n \nmid r$，$r\in \mathbf{Z}$。

求证：不定方程 $x^n+y^n=z^r$ 有整数解。

5．若 $x_i>0$，$\sum_{i=1}^{n} x_i=1$，$x_{n+1}=x_1$ $(n>6)$。

求证：$\prod_{i=1}^{n}(x_i+x_{i+1})^{-1}=n!$。

6．求证：数列$\{\operatorname{tg} n\}$当 $n\to+\infty$ 时是发散的。

7．求证：$\frac{(n+1)^{n+1}}{n^n}>\frac{n^n}{(n-1)^{n-1}}$ $(n=2,3,4,\cdots)$。

（张思明选编，陈剑刚校）

① 这个专栏是由北京大学附中陈剑刚主持下编写的。每期都由具有丰富的中学数学教学经验的教师给出若干有趣味性的数学问题，在下一期给出这些问题的解答。——编者注

## 数学小丛书——智慧之花

## (3)

主 要 目 录

［Ｇｅｎｅｒａｌ Ｉｎｆｏｒｍａｔｉｏｎ］
书名＝邮票·自行车·果园·雨中行 － 丁石孙
作者＝ＢＥＸＰ
ＳＳ号＝
加密地址＝
下载位置＝ｈｔｔｐ：／／ｂｏｏｋ５．５ｒｅａｄ．ｃｏｍ／３００－５５／ｄｉｓｋｊｙ１／ｊｙ１／５８／２５／！００００１．ｐｄｇ

书名
目录
正文